

A MIXED METHODS STUDY EXPLORING THE REALITIES AND PERCEPTIONS
OF PRINCIPAL EVALUATION

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Education

with a

Major in Educational Leadership

in the

Department of Graduate Education

Northwest Nazarene University

by

Donald P. Bingham

April 2013

Major Professor: Paula Kellerer, PhD

AUTHORIZATION TO SUBMIT DISSERTATION

This dissertation of Donald Bingham, submitted for the degree of Doctor of Education with a major in Educational Leadership and titled "A Mixed Methods Study Exploring the Realities and Perceptions of Principal Evaluation," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies.

Major Professor Paula D Kellerer Date 5/1/13
Dr. Paula Kellerer

Committee Members Dr. Burton Webb Date 01 May 2013

Ronald D. Tolman Date Apr. 26-2013
Dr. Ronald Tolman

Department Administrator Loredana Werth Date 5/1/13
Dr. Loredana Werth

Discipline's College Dean Paula D Kellerer Date 5/1/13
Dr. Paula Kellerer

© Copyright by Donald Bingham 2013

All Rights Reserved

ACKNOWLEDGEMENTS

I would like to recognize all of those people who supported me through this inspiring and challenging process. It would be impossible to name all those who helped contribute to this work, but I would like to acknowledge a few that have especially contributed. My wife and children who have shown incredible patience and kindness while I have been working on the research associated with this paper. Dr. Ron Tolman, my mentor and colleague, for helping me and encouraging me to keep working even during those difficult stretches of research and providing flexibility so that I could both work and be a student. Mrs. Yvonne Thurber, my collaboration partner and fellow statistics survivor. Finally, Dr. Paula Kellerer, for her faith, optimism, patience, and vision of what my research could become and who I could become as a researcher. To all of you I thank you. This work would not be possible without each of you.

DEDICATION

This dissertation is dedicated to my father and mother, Randy and Janet Bingham, for their unfailing love and stalwart belief that all children can learn, even when the experts say it is not possible.

ABSTRACT

Principals play a critical role in how a school performs and what type of culture it will have. However, most principal evaluation relies heavily on practices and beliefs grounded in 20th century educational paradigms or adheres to outdated behavioral checklists. This mixed methods study explores the current realities of principal evaluation in southeastern Idaho and the perceptions of 127 principals and superintendents. Findings suggest that there is a variance between written evaluation policies and current practices. Further, there was a statistically significant difference in the perceptions of principals and superintendents in nearly every aspect of principal evaluation studied. The study supports the need to revise and examine current practices being employed to assess principal job performance.

TABLE OF CONTENTS

AUTHORIZATION TO SUBMIT DISSERTATION.....	i
ACKNOWLEDGEMENTS	ii
DEDICATION	iii
ABSTRACT	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
Chapter I Introduction.....	1
Statement of the Problem	3
Background	4
Research Questions	7
Description of Terms.....	7
Significance of the Study	9
Overview of Research Methods	9
Chapter II The Literature Review	11
Introduction	11
Current Status of Principal Evaluation	11
Effective Evaluation Systems.....	15
Conclusion.....	24
Chapter III Design and Methodology	27
Research Design	27
Participants	28
Data Collection.....	29

Analytical Methods	29
Limitations.....	31
Chapter IV Results.....	32
Introduction	32
Summary of the Results	32
Chapter V Conclusion.....	66
Introduction	66
Summary of Results	66
Conclusions	71
Recommendations for Further Research.....	71
Implications for Professional Practice.....	73
Conclusion.....	74
References.....	76
Appendix A.....	87
Appendix B	89
Appendix C	92
Appendix D.....	94
Appendix E	96
Appendix F.....	97
Appendix G.....	99

LIST OF TABLES

Table 1 <i>Template Analysis of Evaluation Policies</i>	33
Table 2 <i>Forms of Evaluation Criteria</i>	35
Table 3 <i>Comparison of Research-based Criteria and District Policies</i>	36
Table 4 <i>Disaggregated Multiple Sources of Data</i>	38
Table 5 <i>Cronbach's Alpha of Questionnaire</i>	40
Table 6 <i>Percentage of Participants Completing the Survey</i>	41
Table 7 <i>Mann-Whitney U Results Superintendent/Principal Comparison – Purposes and Practices</i>	45
Table 8 <i>Mann-Whitney U Results Superintendent/Principal Comparison – Criteria</i>	51
Table 9 <i>Mann-Whitney U Results Superintendent/Principal Comparison – Artifacts and Sources</i>	56
Table 10 <i>Mann-Whitney U Results Superintendent/Principal Comparison – Principal Role in Evaluation</i>	60

LIST OF FIGURES

Figure 1 <i>Questionnaire Respondents</i>	41
Figure 2 <i>How longe have you been in your current position?</i>	43
Figure 3 <i>Superintendent Responses to Items 4a-4d</i>	46
Figure 4 <i>Principal Responses to Items 4a-4d</i>	46
Figure 5 <i>Superintendent Responses to Items 5a-5e</i>	48
Figure 6 <i>Principal Responses to Items 5a-5e</i>	49
Figure 7 <i>Superintendents Responses to Items 6a-6h</i>	52
Figure 8 <i>Principal Responses to Items 6a-6h</i>	53
Figure 9 <i>Comparison of Agreement of Seven Items</i>	57
Figure 10 <i>Comparison of Seven Items which were Statistically Different</i>	58
Figure 11 <i>Comparison of Item 8</i>	59
Figure 12 <i>Comparison of Superintendents and Principals on Items 9a-9d</i>	62

Chapter I

Introduction

In this time of heightened educational accountability and increased accountability for individual school performance, the role a principal plays has never been more important (DuFour, DuFour & Eaker, 2008; DuFour & Marzano, 2011; Reeves, 2004). Along with legislated accountability, there is a growing body of research that has successfully linked strong educational leaders to increased student performance (Brady, 2012; Brown, 2002; Hatrick, 2008; Marzano, Waters, & McNulty, 2005; Raymer, 2006; Rooney, 2005; Taylor, 2011). The most convincing evidence of this correlation was found in the Marzano, Waters, and McNulty meta-analysis of 69 studies. They found a significant positive correlation between school leaders and student achievement. In their research, they identified and defined critical behaviors that were exhibited by principals of high achieving schools. In addition to Marzano et al., Cotton (2003) and Stiggins and Duke (2008) have quantified the effect size of various behaviors that are exhibited by principals that have a positive impact on student achievement. Cotton found that a principals' direct interaction with students had little impact on how student achieve; however, their role as an educational leader in the school had a strong indirect effect on student achievement.

Growing research-based evidence supporting a connection between principal behaviors and actions with student outcomes have resulted in an increased focus on the role and responsibilities of principals. Performance standards such as those set forth by

the Interstate School Leaders Licensure Consortium (Chief State School Officers, 2008) and the nine characteristics of high performing schools (Shannon & Bylsma, 2007) have established criteria to connect principal job performance with behavioral expectations that improve student performance. Likewise, the United States Department of Education has recognized the importance of the school principal and requires the replacement of the principal in two of the four reform methodologies mandated to improve persistently low performing schools (Martineau, 2012). Even though there is an increase in evidence and awareness on the impact of a principal's behavior on a school's achievement and climate, there is very little research on how to accurately measure and assess a principal's job performance (Davis et al., 2011; Studebaker, 2000).

In their study of principal evaluation, Thomas, Holdaway, and Ward (2000) found most superintendents and a majority of the principals in their study felt their evaluation systems accurately reflected a principal's behaviors. Thomas et al. further expressed their belief that evaluations and improved job performance were connected. They go on to outline six major changes and improvements that were required to make the evaluation systems being studied align with their stated belief. These changes included personal reflection, objective feedback, discussion regarding performance, ongoing coaching, and the need for open communication about the evaluation process between principals and supervisors. A similar conclusion was found in Reeves' (2004) work on assessing educational leaders' job performance. He concluded that most educational evaluation systems needed to be revised or replaced due to being outdated, misaligned, and lacking relevance (Danielson & McGreal, 2000; Goldring, Porter, Murphy, Elliot, & Cravens., 2009b; Kearney, 2005).

Statement of the Problem

Jim Collins (2001) wrote, “You must ...have the discipline to confront the most brutal facts of your current reality, whatever they might be” (p. 13). An important question that should be asked has to do with is the current reality for assessing a principal’s job performance. The fundamental problem addressed in this study is comparing what research has found to be effective practices in school leadership evaluation to what is currently being practiced in the field. Research has been able to articulate specific characteristics and behaviors of educational leaders that impact student achievement (Marzano et al., 2005; Stronge, Richard, & Catano, 2008). In addition, there is strong research outlining the specific practices that support effective performance evaluation. In Danielson and McGreal’s (2000) work on teacher evaluation, they found evaluation systems that had the biggest impact on teacher behavior incorporated both a clearly defined criteria to measure job performance and a component focused on individual professional growth. This same finding was echoed in Danielson’s (2007) later work. Thus, it becomes essential to establish the current reality of principal evaluation and identify if current evaluation systems are aligned with research. Further, it is essential to determine if the evaluation systems currently being used measure the behaviors found to increase student achievement and if they are being employed to measure principal job performance. Reeves (2004) calls this the “knowing-doing gap” and states that there is a growing gap between what we know works in educational leadership and how educational leader’s job performance is evaluated.

Research conducted by Goldring et al. (2009a) on principal evaluation systems determined that most systems they studied lacked any congruency between the system

used to assess job performance and behaviors that have been shown to improve school performance. In his study conducted in Israel, Gaziel (2003) found that 75% of the principals felt that the evaluation system used to assess their performance was irrelevant and that it did not adequately assess job performance. Brady (2012), Brooks and Voss (2008), and Conca (2008) found a similar pattern in their studies. Principals perceived that the evaluation systems did not reflect the leadership practices or behaviors that have been proven to impact school performance. Likewise, most principals felt that the process lacked meaning and was not effective in improving a principal's leadership abilities. Conca found that superintendents need additional time and training to adequately and effectively evaluate principals. These three studies found that most evaluation systems lack clearly defined expectations or standards of performance, consistent implementation, personal reflection upon practice, and are not focused on professional growth. In addition to the systematic issues related to administering principal job assessment, there appears to be a consistent divergence between principals and superintendents regarding their perceptions about the purpose and the effects of principal evaluation (Goldring et al., 2009a; Gaziel, 2003; Thomas, Holdaway, & Ward, 2000).

Background

Human-beings have an inherent desire to evaluate their progress towards established goal or expectations. This desire can be either intrinsic or extrinsic, but of late the trend has been greater emphasis on extrinsic acknowledgement than in the past (Wiley, 1997). This has resulted in the evaluation process becoming a critical function in almost every industry throughout the United States (Longenecker & Nykodym, 1996). The evaluation process has evolved into a symbiotic relationship between an employer

and an employee where: the employer is provided information regarding job performance and effectiveness and employees are provided feedback and acknowledgement for their contribution to an organization's progress (Brady, 2012; Longenecker & Nykodym, 1996; McGahie, 1991). Although evaluation has become a central aspect in nearly all work situations, there is evidence to suggest that there is a substantial gap between research on performance evaluations and how evaluations are administered in the field (Arvey & Murphy, 1998).

In looking at evaluation systems in an educational context, the gap found in industry seems to be even more pronounced in the realm of education (Brady, 2012). Evaluation appraisal systems, specifically related to principal evaluation, have remained inconsistent and disconnected from new trends in educational leadership research (Brady, 2012; Heck & Glasman, 1993). In addition to being inconsistent there is very little research focused on what the current realities are and if there is any positive or negative impact on student achievement due to the evaluation process.

Compounding all of this is the developing role of a principal, which makes quantifying and measuring job performance even more difficult (Catano & Stronge, 2007; Marzano et al., 2005; Reeves, 2004). Historically, the role of the principal has been one focused on managing the fiscal and physical operations of a school. However, that role has steadily changed to include a diverse set of skills that range from budgeting to curriculum development (Archer, 2002). Although the role of a principal has shifted from being primarily managerial to focusing more on instructional leadership, there has been little research to determine if the evaluation systems used to monitor principals'

performance have evolved in step with the changes in the job (Brady, 2012; Stufflebeam & Nevo, 1993; Toler, 2006).

In an effort to define the new face of the principalship, professional organizations, researchers, government agencies, and non-profit organizations have endeavored to define and articulate the role of a principal. This has resulted in the creation of standards, frameworks, and characteristics all focused on improving student performance through changing behaviors of educational leaders (Council of Chief State School Officers, 1996; Danielson, 2007; Marzano et al., 2005; Stronge et al., 2008). However, there has been little research linking any of these standards, frameworks, or characteristics to how principals are evaluated (Babo, 2009; Brady, 2012; Derrington & Sharratt, 2008).

There is ample evidence to support the need for strong educational leaders at the helms of our schools, but there is little evidence showing that evaluation impacts principal behavior or what educational leaders do on a daily basis (Brady, 2012; Catano & Stronge, 2007; Kafka, 2009; Nor & Roslan, 2009; Peterson & Deal, 1998; Taylor, 2011). As education rapidly rolls forth into the 21st century and changes to meet the continually developing demands of our students and society, it becomes imperative the methodology and practices used in assessing educational leaders evolves to adequately reflect the current realities of the principalship (Kearney, 2005; New Leaders for New Schools, 2010; Reeves, 2009). It is essential to identify where practice currently is and to make adjustments that are meaningful and centered in research.

Research Questions

Research questions generally focus scholarly work and provide direction for next steps (Creswell, 2008; Marshall & Rossman, 2011). This mixed methods dissertation addresses four main research questions.

1. What are the current practices in evaluating principals in school districts in southeastern Idaho and do they reflect the daily routines of a principal?
2. What criteria, artifacts and sources are being used to measure principal performance and how do they compare to best practices outlined in research?
3. What role does a principal play in principal evaluation?
4. How are the perceptions of principals and superintendents regarding principal evaluation in southeastern Idaho similar and different?

Description of Terms

The following definitions are offered to provide an understanding for key terms used in this study:

Criteria. This refers to the critical research-based behaviors, expectations, or abilities against which individual job performance is measured. Criteria are also referred to as performance standards, or characteristics. In this study, the words are used interchangeably, but they all refer to the specific expectations that are used to measure job performance.

Evaluation. It is the process where an employee's performance is measured against an established criteria and a judgment is made by a supervisor. In this study, it is the process by which a principal's job performance is assessed by a superintendent, assistant superintendent, or other designated individual.

Evaluation System. It is the tools, procedures, policies, and criteria used by a school district to evaluate the job performance of all the principals working in the district.

Framework. A framework is created from a set of commonly defined characteristics which is attached to a standardized set of performance levels.

Principal. The term principal refers to the educational leader at a school who has the authoritative and moral responsibility for leading and directing the operations of the school. For the purpose of this study, the principal is classified as the one being evaluated.

Performance. The term performance in this study refers to job performance. It is what a principal does on a daily basis in the operations of a school.

Professional Inquiry. This is the process of self-reflection and improvement through feedback, mentoring, and intellectual growth. Professional inquiry is focused on the individual identifying areas to improve and establishing a direction for improvement. Professional inquiry is one of the two critical pillars of a solid evaluation system.

Quality Assurance. The guarantee made by an organization to provide the highest quality services by measuring the performance of the various components of the organization is defined as quality assurance. In this study, quality assurance refers to comparing the job performance of principals to what has been found to impact student learning. Quality assurance is one of the two critical pillars of a solid evaluation system.

Superintendent. The term superintendent refers to the individual hired by the Board of Trustees to administer the affairs of a school district. For the purpose of this study, the superintendent is classified as a principal's evaluator and could also refer to an

assistant superintendent, director, etc. The actual title of the evaluator could vary from district to district and would depend on a district's division of duties.

Significance of the Study

Accountability has become a catalyst for most of the reform efforts in education over the past decade. As a result of the increased public and governmental scrutiny of education, it has become increasingly important for a school principal's job performance to be accurately evaluated. One of the potential benefits of this study is to provide more accurate and timely information regarding principal evaluation systems to help state and federal policy-makers as they frame new administrator evaluation law and procedures. Although beneficial to the upper-levels of educational decision making, the primary group that this study is meant to support and provide direction to is at the local district and school level. The findings outlined in this study are targeted to help local superintendents, trustees, and principals develop, rework, and improve principal evaluation systems for the betterment of principal job performance and to improve student achievement.

Overview of Research Methods

This study drew upon tools and strategies from both quantitative and qualitative research using a mixed method approach. The study was composed of three phases; an in-depth literature review, data collection process, and an analysis phase. The first phase focused on synthesizing research and determining best practices in performance evaluation. Also, during the literature review, critical behaviors exhibited by high-performing principals were identified.

Data collection included two collection processes. The first focused on evaluating documents, policies, and practices. The documents collected were analyzed using a template (Appendix A) created by the researcher using information collected during the literature review phase. The researcher employed the qualitative process of template analysis to capture the essential themes and characteristics of current practices used to evaluate principals. The template used in the analysis was made up of five categories, which included evaluation criteria, multiple sources of data, professional inquiry, self-assessment/reflection, and mentoring /collaboration.

In addition to the template analysis, a questionnaire was e-mailed to 71 superintendents and 250 principals in southeastern Idaho. The questionnaire was designed in such a way that all responses were completely anonymous. The purpose of the questionnaire was to establish the current realities of principal evaluation; the role the principal plays in the evaluation process, what criteria is used in principal evaluation, and the perception of whether the current practices are effective.

The final phase focused on analyzing the data using quantitative and qualitative methods. The Mann-Whitney U test was used in the quantitative analysis to determine if there was a difference between principal and superintendent perceptions. Using standard qualitative methods, the researcher analyzed the open ended questions of the survey using content analysis to generate general themes. The analysis included organizing, sorting and coding the responses in like categories and themes.

Chapter II

The Literature Review

Introduction

In his book on assessing educational leaders, Reeves (2004) presents a paradox regarding principals and evaluation systems. In his paradox, he asserts that the greater the responsibilities and authority given to an individual in a school system, the less likely that individual will receive a systematic and constructive evaluation for job performance. This paradox is supported by a general review of research pertaining to principal evaluation.

The following is an overview of the research that has been conducted related to principal evaluation. The literature review will first examine the current status of principal evaluation and analyze what researchers have found regarding the current system. A summary of the research will be presented outlining the behaviors exhibited by principals that have been identified to contribute to improvement in student achievement. Further, research regarding what constitutes an effective evaluation system will be discussed, including an analysis of the two primary functions of evaluation; quality assurance and professional growth.

Current Status of Principal Evaluation

Although there is a growing body of evidence related to the impact principals have on their school community and student learning, there appears to be a gap between what principals do and how that performance is evaluated (Brady, 2012; Heck &

Glasman, 1993). This is largely due to three factors: evaluation systems and assessments being used to measure performance are not rigorous or valid, principals and superintendents perceive principal evaluation differently, and most principal evaluations systems do not align with best practices in evaluation theory (Danielson & McGreal, 2000; Ginsberg & Berry, 1990; Goldring et al., 2009b; Kearney, 2005; New Leaders for New Schools, 2010; Reeves, 2009).

In a report published by New Leaders for New Schools (2010), a non-profit organization focused on improving instructional leadership, it was determined that the vast majority of principal evaluation tools used to measure performance lacked clear performance standards, validity, and implementation. First reported by Hart (1992) and later supported by Goldring et al. (2009b), evaluation systems were found to often rely upon subjective criteria centered on perceptions and not in fact or focused squarely on managerial behaviors rather than instructional behaviors. In addition, evaluation systems often become so complex or so simplistic that they do not truly measure the complexities of a principalship (Catano & Stronge, 2007; Lashway, 2003).

Another failing of both past and current principal evaluation systems is the gap between how principals and superintendents perceive the effectiveness of evaluation tools and processes. Davis and Hensley (1999) found a great deal of diversity throughout the United States on how principals are evaluated. Their findings were related to the differences in perceptions associated with the purpose and usefulness of the evaluation system held by principals and superintendents. Principals viewed their evaluation as something that happened to them, not something that was useful for improving their job performance. They also felt that their evaluations were unduly influenced by external

political factors, such as, parents, board members, and patrons, rather than on daily practices and performance. One of the most cited reasons for this perception was that the person conducting the evaluation based their judgments on limited data. Basing an entire evaluation upon one short visit to the school and little else would be one example of this. Superintendents in that same study felt that their evaluation systems were well developed and were the basis for all performance improvements by their principals. This gap between principals and superintendents perceptions is echoed in Thomas et al.'s (2000) study on policies and practices related to principal evaluation. They found wide-spread diversity between how principals are evaluated and a substantial difference between how principals and superintendents view the importance and usefulness of principal evaluation. Principals felt that they were being evaluated with a narrow focus rather than a holistic view of their performance. They also felt that they did not receive enough support from central office administration in helping them grow and evolve as a principal. Specifically mentioned was the lack of resources at the district level to provide mentoring, frequent feedback, and on-sight visits to principals.

Another common issue in evaluation is the lack of input from more than one source. Traditionally, principal evaluation has been conducted annually by a superintendent or other central office administrators using brief observations of the principal. Researchers often found that these observations were conducted with little or no context or based upon hear-say or sensationalized one-time situations (Andrews, 1990; Davis & Hensley, 1999). Moore (2009) makes the argument that input must be drawn from a wide variety of sources, specifically from those who work with the principal on a daily basis. He asserts that the role of a principal is far too complex to evaluate using a

single perspective or source of information. Reeves (2009) found that 60% of principals he studied felt that their evaluation had no impact on their job performance. Further, 53% stated that they had not received enough specific feedback from their supervisor to make changes to their behaviors that would improve job performance. This was largely due to evaluation systems lacking clear and concise criteria focused on instructional behaviors that improve instructional quality thus impacting student performance (Goldring et al., 2009b; Murphy & Pimental, 1996).

Over the past decade, due to various federal mandates, there has been an increased focus on teacher performance and evaluation. Principal evaluation, however, has remained a low priority for most school districts and states (Kearney, 2005; Hart, 1992; Moore, 2009; Murphy & Pimental, 1996; New Leaders for New Schools, 2010). This is largely due to the increased demands placed upon schools from federal and state mandates related to accountability. Lashway (2003) reported that principals or superintendents had no plans to address principal evaluation, due in large part to the overwhelming number of mandates being placed upon schools and the lack of urgency to make changes to the evaluation system. He postulated that the reason there was not a sense of urgency was due to the major focus in education being on improving student achievement through instructional improvement. He argued that a stronger connection needed to be made between the role of a principal and the effect that role has on student achievement before there would be an increase in the urgency to reform principal evaluation systems. Since 2003, there have been several significant studies that have provided a connection between student achievement and principals (Brown, 2002; Hatrick, 2008; Marzano et al., 2005; Raymer, 2006; Rooney, 2005; Taylor, 2011). Most

notably was the meta-analysis conducted by Marzano et al., who found that there is a positive correlation between principal's leadership skills and a school's student achievement. The study further identified 21 behaviors that constituted the definition of leadership skills. In addition, several studies have found that principals who have effective evaluations with frequent feedback lead schools who tend to have higher levels of student performance (Grissom & Loeb, 2009; Kafka, 2009; Williams, Persaud, & Turner, 2008).

Effective Evaluation Systems

Effective evaluation systems are composed of two primary components: quality assurance and professional inquiry. Danielson and McGreal (2000) examine these two components in their work and assert that every system, regardless of their purpose, educational, sports, business, medical, etc., evaluates its employees based upon these two concepts. Quality assurance is defined as the compulsion an organization feels to make a judgment about how well their employees are performing related to job expectation, thus allowing an organization to guarantee quality assurance to their customers. Danielson and McGreal also contend that it is in the organization's best interest to use the evaluation process to help individual employees grow and flourish in their assigned role. However, according to Danielson and McGreal, before quality assurance can be guaranteed, there must first be a common definition and understanding of what is expected. In their work in 2000 and later in Danielson's work in 2007, they set forth the need for a commonly held and defined criteria or framework to measure performance against.

In addition to quality assurance, Danielson and McGreal assert that strong evaluation systems include employees as part of the process. They categorize this under

the heading of professional inquiry. Professional inquiry means to self-assess and reflect upon practice and establish professional growth goals to improve performance based upon individual need. Danielson and McGreal believed that this would empower individual employees to direct their own professional growth based upon the established criteria used to measure job performance. By incorporating both quality assurance and professional inquiry, evaluation systems are able to achieve a balance between the demand for high quality and the need for ongoing professional growth (Danielson, 2007; Danielson & McGreal, 2000).

Quality Assurance. In principal evaluation, criteria, also referred to as performance standards, are used to measure performance. Danielson and McGreal (2000) and Kearney (2005) determined that in order for an evaluation system to provide quality assurance, it must have a highly defined set of standards to measure performance against. The standards being used to measure principal performance must have both empirical evidence to support the impact on job performance and be linked to what principals actually do on a daily basis (Catano & Stronge, 2007; Council of Chief State School Officers, 1996; Danielson, 2007; Kearney, 2005; Kimball, Milanowski, & McKinney, 2009). In addition to the standards, the system ranks or assigns performance for each standard based upon well-defined performance levels. This creates a framework that can be used in evaluation. It is essential that those who use it be trained and have a common understanding of what each standard and level of performance means in order to establish inter-rater reliability. Without inter-rater reliability, the framework is not reliable and therefore results would not be deemed valid (Creswell, 2008; Council of Chief State

School Officers, 1996; Danielson & McGreal, 2000; Goldring et al., 2009b; Hessel & Holloway, 2002; Tanner, 2012).

One of the problems with standards used in principal evaluation is that they do not meet the established requirements to ensure the quality assurance mentioned above (Harrison, 1988; Harrison & Peterson, 1987; Kearney, 2005; Moore, 2009; Stine, 2001). Condon and Clifford (2010) conducted a study to determine the rigor of principal evaluation assessments throughout the United States. They were able to identify 20 prominently used principal evaluation assessments. After analyzing their results, they determined that only eight of the assessments were reliable and valid in measuring principal performance. In addition to the tools often being unreliable, many principals report not knowing the criteria that is being used to evaluate their performance, which would result in not being able to guarantee quality assurance (Harrison, 1988; Harrison & Peterson, 1987; Kearney, 2005; Stine, 2001). Many educational leaders are falling into the paradox stated by Reeves (2004) in the introduction to this chapter. They are leading their schools, not knowing what is expected and how their job performance is going to be measured.

Principals working in systems that provided a framework for performance and ongoing support and feedback related to job performance had a higher level of job satisfaction. Kimball, Milanowski, and McKinney (2009) found that principals working in this type of a system reported that they received better feedback compared to a group that was evaluated using traditional approaches. They also reported that they felt empowered to take control of their professional learning, and perceived themselves as instructional leaders rather than managers. In addition, principals who were able to self-

assess their performance against standards and participated in mentoring programs tended to have schools with positive cultures and higher student achievement (Bulach, Boothe, & Pickett, 2006; Grissom & Loeb, 2009; Hall, 2008; Prestine, 2008; Robinson, Eddy, & Irving, 2006).

Effective criteria. Through a synthesis of research a long list of critical areas needed to be addressed in a principal's evaluation criteria emerges. Catano and Stronge (2007) found that strong evaluation criteria should be aligned to state and professional standards and also focus on instructional leadership, organizational management, and community relationships. William, Persaud, and Turner (2008) found that principal evaluation must focus on interpersonal skills as these skills tended to significantly impact student performance. Marzano, Waters, and McNulty (2005) outlined 21 responsibilities of a principal that were essential in improving student achievement at the school level. These responsibilities included: affirmation, change agent, contingent rewards, communication, culture, discipline, flexibility, focus, ideals/beliefs, input, intellectual stimulation, involvement in curriculum, instruction, and assessment, knowledge of curriculum, instruction, and assessment, monitoring/evaluating, optimizer, order, outreach, relationships, resources, situational awareness, and visibility. Earlier research conducted by Marzano (2003), outlines five critical factors that principals directly impact which complement the 21 responsibilities. These factors include: guaranteed and viable curriculum, challenging goals and effective feedback, parent and community involvement, safe and orderly environment, and collegiality and professionalism. Regardless of the criteria used, Kearney (2005) makes a strong case for the necessity of whatever criteria is to be used; it must be grounded in the realities of the job.

In reviewing the literature on principal evaluation, two specific pieces of research are most often cited. The first is the definitive research on establishing the effective traits of successful leaders conducted by Marzano et al. (2005). In their research, a meta-analysis was used to determine what characteristics of principals had the largest impact on student achievement. The results of their study created a list of 21 characteristics that strong school leaders exhibited. Stronge, Richard, and Catano (2008) built upon their research and incorporated national standards into their criteria. They narrowed the categories into a more manageable set of eight qualities. Those qualities are instructional leadership, school climate, human resource administration, assessing instructional quality, organizational management, communication and community relations, professionalism, and student achievement. In addition to the categories, they also defined performance levels related to the categories, which was something that was not found in Marzano et al.'s work. Those levels include, master, professional, apprentice, and ineffective. For the purpose of this study, the work of Stronge et al. will be used as the criteria as it is both aligned to standards and best practices grounded in research.

Inter-rater reliability. The other critical aspect of quality assurance is ensuring evaluators are accurate, consistent and base their judgments upon evidence, not opinion (Danielson & McGreal, 2000). In order for any evaluation system that is based on observation and the collection of evidence to be valid, it must be able to demonstrate inter-rater reliability. Creswell (2008) defines inter-rater reliability as the ability of two or more observers ability to score a behavior similarly. Research has identified that effective evaluators should be able to recognize examples of the evaluative criteria in action, interpret evidence against the evaluative criteria, make judgments about the

performance, and consistently be able to link those judgments to the descriptions of levels of performance (Danielson, 2007; Danielson & McGreal, 2000). In order to do this, evaluators must participate in comprehensive professional development which incorporates modeling, coaching, collaborative conversations, and field experience (Davis & Hensley, 1999; Ediger, 1998; Hall, 2008; Stine, 2001; Thomas et al., 2000; Walker & Qian, 2006). Principals who are evaluated using a standards-based criteria and who receive feedback and mentoring tend to exhibit stronger leadership skills that are associated with successful schools (Bickman et al., 2012; Hall, 2008; Prestine, 2008; The Education Alliance at Brown University, 2004).

Professional Inquiry. In addition to providing quality assurance to the public and profession, evaluation systems should also develop and promote professional inquiry. Professional inquiry can best be described using the following attributes: reflection on practice, collaboration with peers, self-assessment of performance, and professional development (Danielson, 2007; Danielson & McGreal, 2000; Davis & Hensley, 1999; Thomas et al., 2000).

Reflection on practice. A critical practice in professional practice is that of self-reflection. Meaningful and systematic reflection on job performance is one of the most powerful components of any evaluation system (Blum, Butler, & Olsen, 1987; Bush & Chew, 1999; Danielson, 2007; Danielson & McGreal, 2000; Knight, 2007; Thomas et al., 2000). Part of human nature is to reflect upon performance in almost any circumstance and identify areas of strength and areas for improvement.

Reflection on practice is more than just making self-judgments. It is about critically analyzing performance against the established job expectations or criteria,

celebrating success and growth, and charting a plan for improvement. Danielson (2007) gave this counsel,

To be productive, reflection on practice must be systematic and analytic. When a lesson has not gone well, it is important for a teacher not only to recognize that it was not successful but also to be able to determine the reasons for that outcome. Only if those reasons are understood can a teacher improve the lesson the next time. (p. 169)

This is as applicable to principals as it is to teachers. This is where the role of a coach or mentor can be extremely beneficial. A superintendent or veteran principal can often provide valuable insight to a principals as they reflect on their performance and their school's performance (Bickman et al., 2012; Blum et al., 1987; Ediger, 1998; Gaziel, 2003; Ibukun et al., 2011; The Education Alliance at Brown University, 2004). Although there are relatively few studies regarding principal mentoring, the initial findings are encouraging (Knight, 2007; Prestine, 2008). In a study comparing the model of principal mentoring in the United Kingdom to that found in Singapore, Bush and Chew (1999) found that, although the two styles of mentoring were different, they both had the following effects: improved performance of the new administrator, reinvigorated veteran administrators who were assigned as coaches, and improvement in the educational systems as a whole through helping new administrators become proficient more quickly and grounded in a school system's culture and climate. Likewise, Knight postulates that coaching is a far more effective way of implementing change and improving instruction than any other professional development technique. This same

assertion is made by DuFour, DuFour, and Eaker (2008) who cite learning as a professional team as the best method for improving student learning and school culture.

Peer collaboration. Peer collaboration is closely tied to mentoring and coaching and uses collegial support as a means to improve and enhance job performance. Often these collegial teams are called professional learning communities. The power of professional learning communities on improving student achievement and the positive impact of peer collaboration on school climate has been greatly discussed throughout the literature (DuFour et al., 2008; DuFour & Eaker, 1998; Eaker, DuFour, & DuFour, 2002; Feldman & Ouimette, 2004). One of the defining hallmarks of professional learning communities is the focus on collective learning through community inquiry. In the work of the DuFour et al., they demonstrate how groups of educators organized into meaningful teams create powerful learning communities that promote professional inquiry and student learning.

In the area of the principalship, it can be difficult to collaborate on administrative issues and concerns as most principals work in isolation. This is especially true of principals working in rural areas. Fahey (2011) found that principals who participate in professional consultancy with other principals were able to negotiate the complexities of the principalship. Further, principals report that their understanding and skills as a leader were deeper through collective learning. Similar results have been found in professional learning communities comprised of principals (Blum et al., 1987; Brody et al., 2010; Mitchell & Castle, 2005; Walker & Qian, 2006). In Stine's (2001) research, he affirms this notion that effective evaluation should include a collaborative system that brings a principal and supervisor together to develop avenues for improved performance.

Self-assessment of practice. Many traditional evaluation systems focus on one view of a principal's performance using a snap-shot in time. A new trend is beginning to surface which is based on the concept of self-assessment or self-evaluation. The critical need for principals to play a key role in their own evaluation and the evaluation process as a whole was first mentioned by Harrison and Peterson (1987) and later supported by Davis and Hensley (1999). Bulach, Boothe, and Pickett (2006) also found that principals who take an active role in their evaluation through self-assessment and reflection linked to a criteria have a greater impact on student performance and school culture.

Deeper understanding and commitment can be found in those who self-assess their performance (Andrews, 1990; Danielson & McGreal, 2000; Schrader & Steiner, 1996). Further, evaluation is strengthened and more meaningful when both the evaluator and the person being evaluated have the same understanding regarding that individual's performance (Danielson, 2007; Danielson & McGreal, 2000; Schrader & Steiner, 1996).

Professional development. The final component of professional inquiry is professional development that is based upon individual needs related to the established job criteria (Blum et al., 1987; Danielson & McGreal, 2000; Marcoux, Brown, Irby, & Lara-Alecio, 2003; Sun & Youngs, 2009). Professional development is the practice of identifying an area of need and seeking out knowledge and best practices that have been proven to work. Traditionally, the professional growth of employees has been nonexistent or composed of various workshops that are not specifically connected to areas needing improvement. The result of this approach is gleaning only what is desired and often the new knowledge is not implemented (DuFour et al., 2008; DuFour & Eaker,

1998; Reeves, 2006). However, with the increased focus linking student performance to principal job performance, professional development has become increasingly important.

Evaluations tend to be disconnected from the realities that principals face on a daily basis (Danielson & McGreal, 2000). As a result, there is a perception that it is just a hoop that must be jumped through annually. When a principal or teacher has the opportunity to self-reflect on their practice, discuss their practice with colleagues, self-assess their performance against a legitimate criteria, and make a plan of action to grow as a professional, evaluation is viewed more seriously (Marcoux et al., 2003; Sun & Youngs, 2009). It is also important to understand that when professional development is linked to specific areas of an evaluation, the professional development experience takes on more meaning and value. If a principal has self-reflected and assessed their performance and collaboratively discussed it with the superintendent, a whole new world of professional development opportunities arise that are meaningful to the principal and linked to performance (Marzano & Waters, 2009; Sun & Youngs, 2009).

Conclusion

The evaluation process should be a continuously evolving and improving experience. However, in education it often gets relegated to the back of a long line of initiatives and mandates. Often times, educational evaluation systems are outdated, limited, or lack creditable and valid criteria to measure performance against (Condon & Clifford, 2010; Danielson & McGreal, 2000; Stine, 2001). With the increased demands and political scrutiny being placed upon schools to improve student achievement, the process used to evaluate principals is becoming increasingly important. With the inception of pay-for-performance, value added evaluation, parent input, and increased

federal and state accountability, the need to understand and implement a valid and rigorous evaluation system is becoming paramount.

As education moves forward, it is important to remember three general assumptions about any principal evaluation system. First, the system must be fair and equitable (Danielson, 2007; Danielson & McGreal, 2000). Principals need to be able to feel that they are being evaluated by someone who knows and understands the role of a principal and that bases their judgments and recommendations upon facts collected from practice (Brady, 2012). Principals also must know that their evaluator is consistent and reliable in how they evaluate all their subordinates (Brady, 2012; Heck & Glasman, 1993). The criteria and levels of performance should be clearly accessible and articulated to principals so that they have a clear understanding of what is expected and how they are going to be evaluated (Danielson, 2007; Reeves, 2004). Secondly, the evaluation should be based upon what principals are expected to do. Principals should be able to look at the criteria of the evaluation tool and match it to their daily practice. Criteria should be specific and clearly defined (Harrison, 1988; Harrison & Peterson, 1987; Kearney, 2005; Stine, 2001). Finally, evaluations should be grounded in research and reality. The evaluation criteria and process should align to best practices outlined in research and standards, but also reflect what is actually happening in practice (Brady, 2012; Catano & Stronge, 2007; Kafka, 2009; Nor & Roslan, 2009; Peterson & Deal, 1998; Taylor, 2011).

The role of the principal is complex, demanding, stressful, and highly rewarding. It is vital that evaluation systems used to assess their performance are not overly complicated or simplified, but align to the job that they perform each day in helping

improve and sustain schools. It is critical that research be conducted to monitor what systems are currently being used and continue to define best practices.

Chapter III

Design and Methodology

The primary purpose of this study was to determine if principals are being evaluated as prescribed by effective evaluation practices outlined in research. Further, the study explored how principals and superintendents perceived principal evaluation and determined if there was a difference between the two. The final and critical component of this study was to determine what was needed to make principal evaluation effective and meaningful to principals. This chapter will discuss the research design, participants, data collection processes, analytical procedures used, and the limitations of this study.

Research Design

Creswell (2008) cautioned researchers using a mix methods design to carefully determine if the approach meets six characteristics. Those characteristics included a rationale for the design, quantitative and qualitative forms of data to be collected, priority, sequence, data analysis matched to a design, and diagram of the procedures. This and subsequent sections of the chapter will address those characteristics. The simplest rationale for using this method was the fact that performance evaluation attempts to capture human behavior and measure those behaviors against an established criteria. A mixed method approach, utilizing both quantitative and qualitative methods, seemed best fitted to glean the most information and data.

The study was composed of three phases which were undertaken over a year long period comprising the 2012-2013 school year. The first phase focused on evaluation of documents,

policies, and practices. Superintendents in 55 school districts, located geographically in the southeastern corner of Idaho, were sent a letter requesting a copy of their districts current policy regarding principal evaluation. In addition to the policy, the letter also requested a copy of their current principal evaluation tool and any written evaluation procedures that were used in evaluating principals. Superintendents were asked to send their documents electronically to the researcher.

The second phase of the study dealt with collecting anonymous questionnaire responses from principals and superintendents in the same 55 school districts and geographical area of southeastern Idaho that was used in the first phase of the study. Principals and superintendents were sent an e-mail using Qualtrics online survey system and asked to complete an anonymous questionnaire.

The final phase was the analysis of the data retrieved from the questionnaires and the document review that will be covered in greater detail later in this chapter.

Participants

The participants in this study were superintendents and principals of all levels from southeastern Idaho. Participants included 301 individuals, including 246 principals and 55 superintendents. The principals included 132 elementary principals, 44 middle school principals, 51 high school principals, and 19 alternative school principals. The researcher used the Idaho State Department of Education Educational Directory which can be found at the Idaho State Department of Education's website to identify the participants. Using the directory, the researcher utilized the state's established regions to create a geographical area for the study. The regions selected were 4, 5, and 6 which comprise the bottom southeastern corner of the State of Idaho.

Data Collection

Prior to data being collected, approval was gained from Northwest Nazarene University's Human Research Review Committee (HRRC). In addition to university oversight and permission to conduct the study, informed consent was obtained from all participants in the study as part of the questionnaire (see Appendix B). Pseudonyms were used during the document analysis process to keep information confidential and help in the data collection process.

Data was collected in two phases: document collection and questionnaires. During the document collection phase, evaluation policies, evaluation tools, and templates were collected through direct request of individual districts or through the internet. In the event that documents were not sent by the districts to the researcher, information was retrieved via the internet using individual district webpages.

An electronic questionnaire was devised by the researcher to collect information related to principal evaluation and evaluation processes. An invitation to participate was sent to all principals and superintendents in the defined study area and data was collected electronically via Qualtrics Online Survey Software. The questionnaire contained 13 items, which included a mix of open ended responses and Likert-scale questions. A copy of the questionnaire can be found in Appendix B.

Analytical Methods

Different analytical methods were used for each phase of the study. Documents that were collected during the first phase were analyzed using the qualitative process of template analysis as outlined by Marshall and Rossman (2011). The process began with the creation of a template of categories and themes. (See Appendix A). These categories and themes were

derived from the literature review based upon effective practices related to principal evaluation. The documents collected were coded using the categories template. Although the categories and themes were established prior to the document analysis, the process allowed for the categories to evolve and change during the analysis and coding process. This allowed new categories to emerge from the documents collected.

Prior to administering the questionnaire, the researcher conducted reliability and validity tests on the questionnaire. The researcher conducted a pilot test of the questionnaire with a sample of nine experts composed of principals and superintendents to establish reliability. The primary purpose of the pilot test was to determine if the survey instrument could be completed by the participants and to collect feedback about each item (Creswell, 2008). In addition to reliability, the researcher also established validity using content validity. The same group of experts that took the pilot test were used to determine the content validity of the questionnaire. The experts were asked to rank each item on a four point Likert-scale as to whether the item was relevant or not relevant to principal evaluation (Lynn, 1986; Polit & Beck, 2006). Further, the experts were asked if the questionnaire as a whole was relevant to the topic of principal evaluation. They were also asked to comment on the items and offer suggestions to improve the questionnaire reliability and validity. In addition to the pilot test and content validity, Cronbach's alpha was used to calculate the internal consistency reliability of the questionnaire.

The information collected from the questionnaire was analyzed using the Mann-Whitney U to determine if there was a difference between the responses of the principals and superintendents. The researcher used a $p=0.05$ to determine if there was a statistical significance between the two groups (Tanner, 2012). The open-ended responses were analyzed using standard qualitative analytical procedures as outlined by Marshall and Rossman (2011). This

included organizing the data, sorting the data, coding the data, generating categories and themes, offering interpretations through analytical memos, searching for alternative understanding, and reporting findings.

Limitations

As with any research there were limitations. This study was no exception. Marshall and Rossman (2011) define limitations as a reminder to readers “of what the study is and is not – its boundaries and how its results can and cannot contribute to understanding” (p. 76). Limitations are considered those influences that cannot be controlled. Delimitations are limitations purposefully used by the researcher to establish boundaries of the study (Marshall & Rossman, 2011).

The biggest limitation of this study was the use of self-reported information through a questionnaire. Although professional and ethical honesty was assumed, there was the possibility of misinformation being reported. Further, there was the possibility of respondents misunderstanding the questions used or unintentionally reporting incorrect information.

An additional limitation was the rural nature of the study area. Southeastern Idaho is mostly made up of small to mid-sized school districts and communities. Although the area has several larger cities, they do not reflect what is considered a large urban area. A potential limitation of this study was its ability to be generalized to systems that are larger and more complex in nature.

Chapter IV

Results

Introduction

This study's primary focus was to establish the current practices, criteria, and sources used to measure principal job performance and to measure principal and superintendent perceptions of principal evaluation in eastern and south central Idaho. This chapter outlines the results of the study. The researcher chose a chronological order to present the results of the study based upon the sequence used to conduct the study. The results are organized by the phases used in the study.

Summary of the Results

Phase one: Template analysis. The first phase of the study focused on the collection of current principal evaluation documents and tools. The researcher contacted 55 school districts requesting evaluation policies and forms. In response to the request, 32 districts responded. This resulted in a 58% response rate to the first phase of the study. Out of the 32 districts that responded to the request, 30 of the districts provided their general policy and additional documentation by way of evaluation procedures or forms. Two districts stated that they did not have an evaluation policy or procedures in place for principal evaluation. They stated that their district used a written narrative outlining job performance with no set criteria or procedures.

The documents provided were analyzed using template analysis (Marshall & Rossman, 2011). Based upon the literature review, the template used initially in the

process included five categories: (a) Evaluation Criteria, (b) Multiple Sources of Data, (c) Professional Inquiry, (d) Self-Assessment/Reflection, (e) Mentoring/Collaborating. The template evolved during the analysis process to include the addition of general policies and no policies or procedures. The following table outlines the overall results of the template analysis:

Table 1

Template Analysis of Evaluation Policies

	Districts
General Policies	30
Evaluation Criteria Outlined	24
Sources of Data Identified	15
Professional Inquiry Described	3
Self-Assessment Included	4
Mentoring or Collaboration Outlined	3
No Policy	2

Source: Template Analysis of Evaluation Policies (Appendix B)

All but two of the districts that responded to the request for documents had some kind of general policy. Six of the district general policies did not define evaluation procedure or criteria as part of their general policies or procedures. During the initial analysis the researcher observed that many of the policies were identical. The policies were separated into matching piles. Upon further investigation, the researcher learned that many districts subscribe to a policy service through the Idaho School Boards Association or through the law firm of Eberharter-Maki and Tappen. Eight of the districts that responded used the template from the Idaho School Boards Association and

nine districts used the policy template from Eberharther-Maki and Tappen. The remaining 13 districts had policies that were unique from the others submitted.

The areas of professional inquiry, self-assessment, and mentoring/collaboration were straightforward and did not require disaggregation. There were four districts whose policies required or included self-assessment as part of the evaluation process. Likewise, there were only three district policies that included some kind of professional development component in their evaluation policy or procedure. This was described in all three policies as developing a professional development or growth plan with goals and action steps related to developing as a principal. The same three district policies also encouraged participation in professional organizations and collegial collaboration which the researcher interpreted as meaning mentoring or collaboration. However, none of the districts in the study had a defined mentoring or peer assistance program outlined in policy or procedure.

While conducting the template analysis, four categories emerged from the documents dealing specifically with the form that evaluation criteria took. For example, were the criteria in the form of a check list or attached to performance levels. The forms included: defined performance levels, Likert-scale, artifacts, and narrative. The following table provides a disaggregated view of the evaluation criteria:

Table 2

Forms of Evaluation Criteria

Identified Form	Number of Districts
Defined Performance Levels	3
Likert-Scale	7
Artifacts	2
Narrative	12
No Description Provided	12

Source: Forms of Evaluation Criteria (Appendix C)

Three of the districts provided a tool that clearly defined the criteria used to measure job performance and a rubric defining each criterion by performance level. Seven of the districts used evaluation instruments that used varying Likert-scales of checklists. Two districts had clearly defined criteria and required evidence by way of artifacts representing actual performance against the defined criteria. However, these districts did not have clearly defined performance levels. There were 12 districts that utilized a narrative piece in measuring the established criteria. The narrative aspects of the evaluation are composed by the evaluator based upon collected data, perceptions, and self-reflection of the principal. Twelve districts that provided policies identifying evaluation criteria did not provide additional materials to analyze to determine how the criteria were used.

Research-based criteria that emerged from the literature review of the study was used to compare the criteria outlined in district policies (Table 3). These themes included: instructional leadership, school climate, staff development, staff evaluation, administration and management, community relations, professionalism, and efforts

towards improving student achievement. Of the 12 districts that provided their criteria and tools for review, 11 used criteria that could be aligned to the themes that emerged in the literature review of this study (Table 3). In all but one of the areas, Management, at least 6 of the districts used research-based criteria to measure job performance.

Table 3

Comparison of Research-based Criteria and District Policies

Criteria	Number of Districts
Instructional Leadership	8
Climate	10
Staff Development	6
Staff Evaluation	8
Management	5
Community Relations	9
Student Achievement	11

Source: Comparison of Research-based Criteria and District Policies (Appendix D)

All the district policies outlined multiple sources of information for an evaluation. This is typified by the use of the following language in the policy, “Such evaluation shall be based on the job description, accomplishment of annual goals and performance objectives, and established criteria.” However, this study took a more narrow view of multiple sources of data, specifically looking for sources of data that were external to the superintendent or evaluator. Using this definition, 15 districts used multiple sources of data. Seven districts incorporated student achievement as part of the evaluation. All 15 districts that reported using multiple sources of data used various forms of parent input. Three districts provided specific surveys that were used to collect parent input. Others

used more general approaches to collect parent input such as verbal input received by principals throughout the year, and e-mails from parents regarding instructional issues both positive and negative. In seven cases, student achievement was used as one source of data which composed 50% of the evaluation (Table 4). In four of those districts, student achievement composed 50% and the other 50% was based upon the evaluation criteria. The remaining three districts used 50% of their evaluation coming from student achievement with various configurations of parent input, staff perceptions, self-assessment, and evaluation criteria as sources of data to achieve the desired 100%.

Eight districts did not include student achievement as part of the evaluation process. They instead used parent input, growth plans, self-assessment, staff perceptions, and evaluation criteria.

While examining the documents, the researcher identified six different sources used by evaluators to gauge performance against the criteria. Those categories included student achievement, parent input, growth plans, self-assessment, staff perceptions, and evaluation criteria. There were 15 districts that included performance criteria as part of their general policies or supporting documents sources of data used for the evaluation.

Table 4

Disaggregated Multiple Sources of Data

Sources of Data	Number of Districts
Student Achievement	7
Parent Input	15
Growth Plans	1
Self-Assessment	4
Staff Perceptions	3
Evaluation Criteria	15

Source: Disaggregated Multiple Sources of Data (Appendix E)

Phase two: Questionnaire.

Reliability and validity. The second phase focused on collecting data through a questionnaire devised by the researcher. The first step in this process was establishing the reliability and validity of a questionnaire focused on the research questions of this study. An important aspect of any study is establishing the reliability of the instrument used. Reliability, as defined by Creswell (2008), “means that the scores from an instrument are stable and consistent” (p. 169). Likewise, an instrument must be valid or in other words make sense and are representative and meaningful (Creswell, 2008). The researcher used both a pilot test and Cronbach’s alpha to determine the reliability and validity of the questionnaire. The pilot test used nine experts that were either principals or superintendents outside of the designated study area. The pilot test served a dual purpose, establishing reliability and also content validity. A content validity index (CVI) was used to calculate validity as outlined by Polit and Beck (2006) and a complete table of the results can be found in Appendix F and G.

Based upon the input from the experts who participated in the pilot test, two items were removed from the questionnaire. The first item dealt with pay for performance being calculated based upon principal evaluation. The second item focused on the use of honors and awards a school received as part of the sources of information that evaluation is based upon. Both items were removed due to the experts feeling that the items were not relative to the topic. This was confirmed by the item's content validity scores, each item received a 0.44, which is below the threshold of 0.78 established by Lynn (1986) for nine experts for an item to be valid. There were two other items that received a CVI score of 0.67 which dealt with student surveys and portfolios being used in principal evaluation. However, based upon the literature review, the researcher felt that it was important to include both of those items in the questionnaire. The overall scale content validity index score of the questionnaire was 0.90 before excluding any items, which is above the threshold of 0.78. After removing the two items that scored an index score of 0.44 the overall scale content validity score of the questionnaire was 0.92.

Cornbach's alpha was also conducted on the questionnaire to establish internal consistency of the questionnaire. This was done for the questionnaire as a whole and also in four designated categories which included: (a) Purposes and Practices, (b) Criteria, (c) Artifacts and Sources, (d) Policy Development. The following table outlines the overall alpha of the questionnaire and each category.

Table 5

Cronbach's Alpha of Questionnaire

Questionnaire/Categories	N of Items	N of Respondents	Cronbach's Alpha
Purposes and Practices	9	123	0.901
Criteria	8	118	0.909
Artifacts and Sources	15	118	0.841
Policy Development	5	121	0.802
Overall Questionnaire	37	113	0.867

According to George and Mallery (2003), internal consistency above 0.80 falls into a good range and internal consistency above 0.90 is in an excellent range. The overall internal consistency as measured by Cronbach's alpha for this instrument is in the good to excellent range, thus establishing it as being reliable.

Response rate and demographic results. The questionnaire was distributed via e-mail to all principals and superintendents that were listed by the Idaho State Department of Education in their directory which was retrieved from their web-site on September 20, 2012. An invitation to participate in the study was extended to 246 principals and 55 superintendents for a total of 301. The questionnaire was open from November 13, 2012 through December 17, 2012. Two reminders were sent to principals and superintendents during this window. At the close of the window, 127 administrators had responded to the survey, 26 superintendents, 3 assistant superintendents, 92 principals, 2 vice principals and 4 others. Two of the four who responded as other were in a combined principal/superintendent position, one was a charter school principal, and the other was a combination principal and federal program director. The researcher combined

respondents who marked that their current position was either superintendent or assistant superintendent into one group called superintendents, classifying them as those that evaluate principals. The researcher combined those responding as principals, vice principals, or other, into a group called principals or those being evaluated (Figure 1). Table 6 represented the response rates by combined categories and total response rate for the questionnaire.

Figure 1

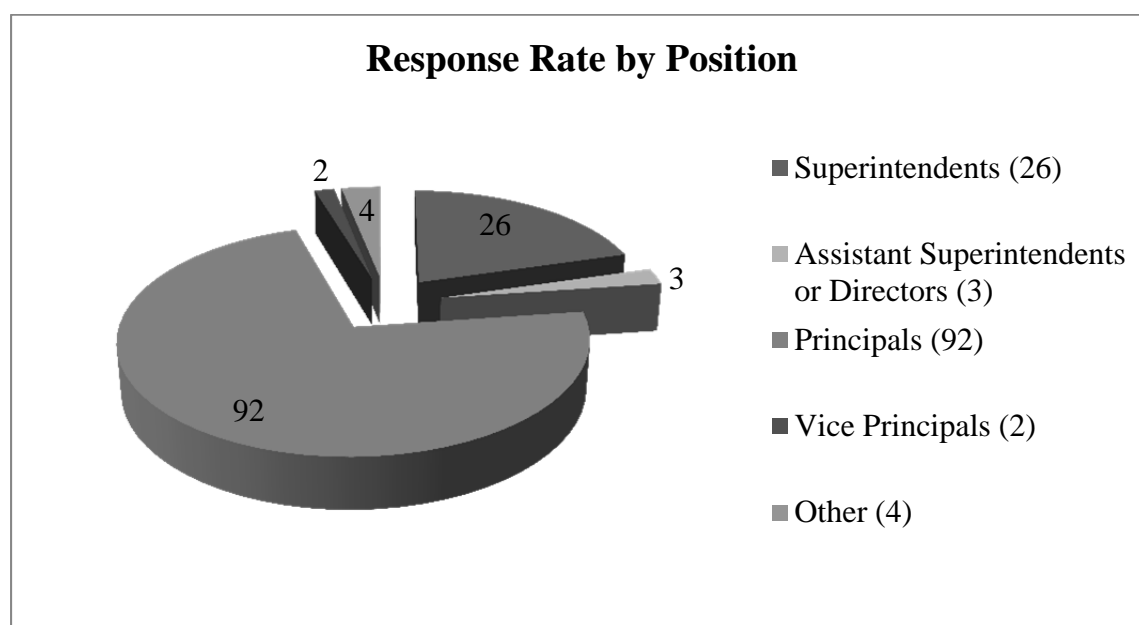
Questionnaire Respondents

Table 6

Percentage of Participants Completing the Survey

Participants	Total Invited	Total Responses	%
Superintendents	55	29	53%
Principals	246	98	40%
Total	301	127	42%

Creswell (2008) stresses the importance of having a strong response rate so that a researcher can have confidence when generalizing results. He cites that most educational journals report a response rate of 50% or better, however, he indicates that this rate will vary when using a survey or questionnaire that is mailed or e-mailed to participants.

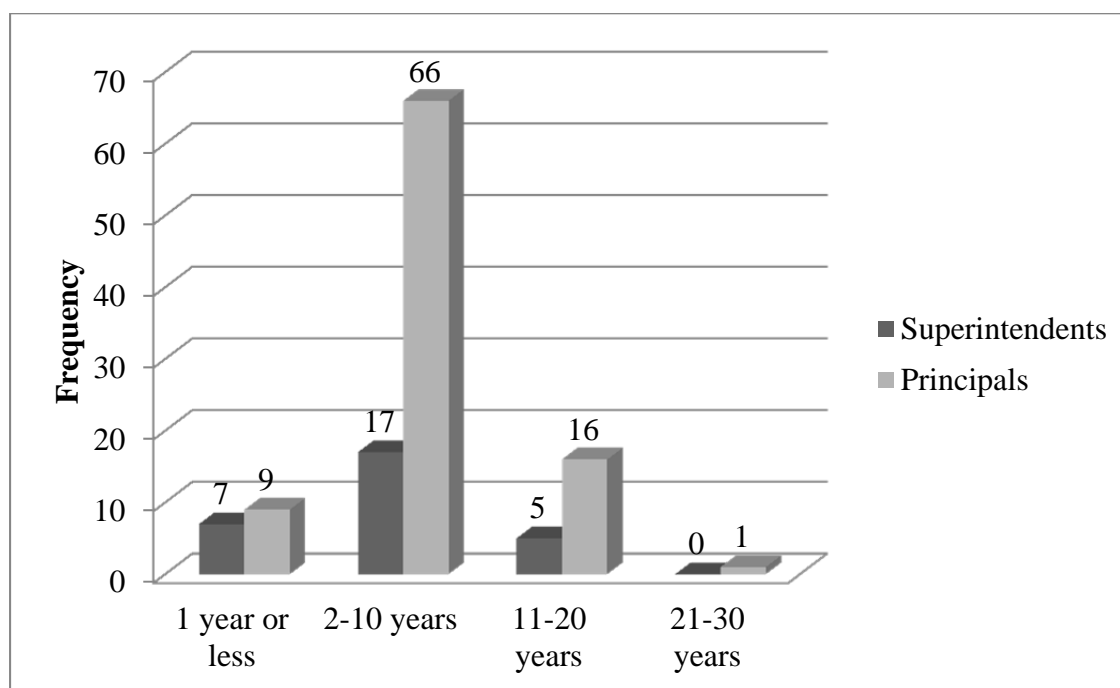
Baruch (1999) conducted a comparative analysis of 141 journal articles which included 175 studies and determined that for most academic studies the response rate should fall above 40%. Using this as a threshold, the response rate was acceptable.

Quantitative Results. One of the primary purposes of this study was to determine the current practices and purposes of principal evaluation. In addition, the study sought to determine if there is a significant difference between the perceptions of principals and superintendents regarding evaluation. The questionnaire was developed around these two purposes and the results of the template analysis and was composed of six sections. Those sections included; demographics, purposes and practices, criteria, artifacts and sources, policy development, and qualitative questions.

Demographics. Limited demographic information was requested from the participants. This was due, in large part, to the rural nature of the geographical area being studied and securing the participants anonymity. Participants were asked to identify what their current position was in their school district (Table 5) and how many years they had been working in that position. Figure 2 shows the distribution of longevity of participants by their classified group.

Figure 2

How long have you been in your current position?



A majority of those that participated in this study, 87%, could be classified as being experienced having more than 2 years in their current position. Specifically, 90% of the principals had at least 2 years of experience in their role as a principal and 76% of the superintendents.

Participants were also asked in the demographic section to identify who in their district evaluated principals. Sixty-five percent responded that the superintendent was responsible for conducting principal evaluation in their school district and 24% indicated that another central office administrator was responsible. Eight percent reported that their evaluation was a self-evaluation. A small percentage of those who responded, 3%, indicated that they had not been evaluated or did not know who evaluated principals in their district.

Purposes and practices. This section of the questionnaire dealt with the purposes of principal evaluation and what practices are currently used in principal evaluation. Participants were given a series of nine statements and asked to complete a Likert scale indicating their level of agreement on a five-point scale ranging from strongly disagree to strongly agree. The Mann-Whitney U test was used to analyze the results comparing the two independent groups; superintendents and principals. The researcher used a $p=0.05$ which would mean that any z value that was ± 1.96 would be statistically significant (Tanner, 2012). Using this methodology, the results of the Mann-Whitney U for this section indicated that there was a significant difference between the perceptions of principals and superintends regarding the purpose and practices of evaluation in all areas, with the exception of using evaluation to determine promotion or tenure (Table 7).

Table 7

Mann-Whitney U Results Superintendent/Principal Comparison – Purposes and Practices

Items	Mann-Whitney U	z	Asymp. Sig. (2-tailed)
4a. In my district principal evaluation reflects what principals do on a daily basis.	868.0	-3.538	0.001
4b. In my district the criteria used in principal evaluation are researched-based.	1008.0	-2.462	0.019
4c. In my district principal evaluation improves principal job performance.	978.5	-2.639	0.005
4d. In my district principal evaluation is effective.	967.0	-2.637	0.005
5a. The purpose of principal evaluation in my district is to promote professional growth as a principal.	1017.0	-2.552	0.011
5b. The purpose of principal evaluation in my district is to assess job performance as measured by a set of clearly communicated criteria or standards.	859.0	-3.287	0.001
5c. The purpose of principal evaluation in my district is to determine promotion or tenure.	1335.0	-0.141	0.965
5e. The purpose of principal evaluation in my district is to improve student achievement.	714.5	-4.165	0.000
5f. The purpose of principal evaluation in my district is to improve instruction in school.	651.0	-4.468	0.000

Notes: If $z \pm 1.96$, it's statistically significant at $p = 0.05$.

Figures 3 and 4 show the responses of the superintendents and principals respectively. Overall, superintendents as a group tended to respond more affirmatively

on the items in this section than did the principals. Principal responses were more varied and on a number of the items a large number of the principals marked uncertain.

Figure 3

Superintendent Responses to Items 4a-4d

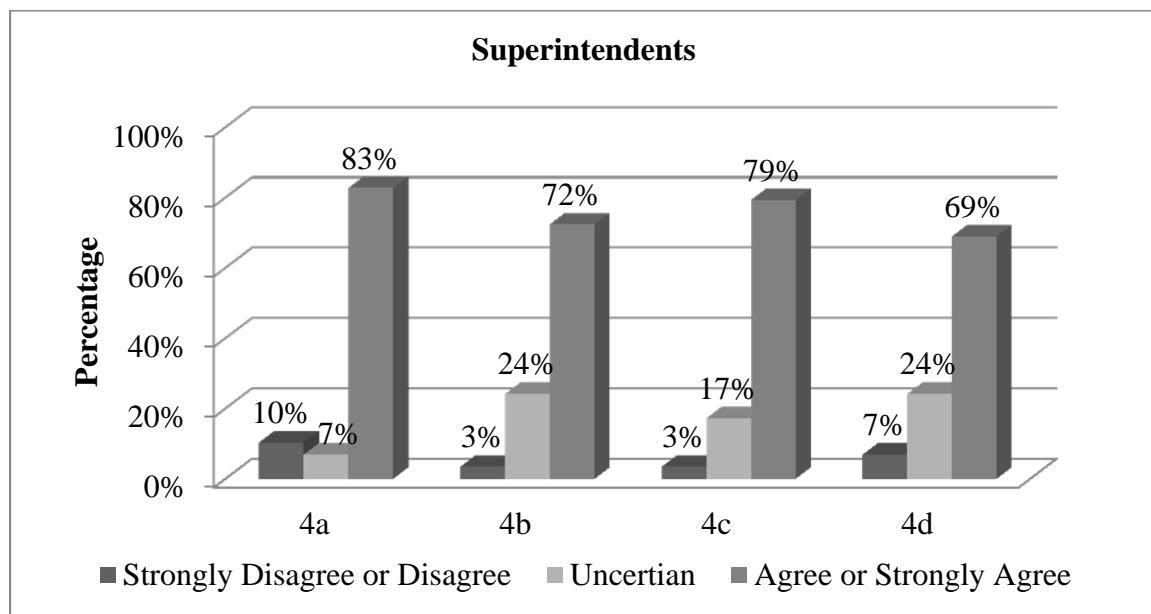
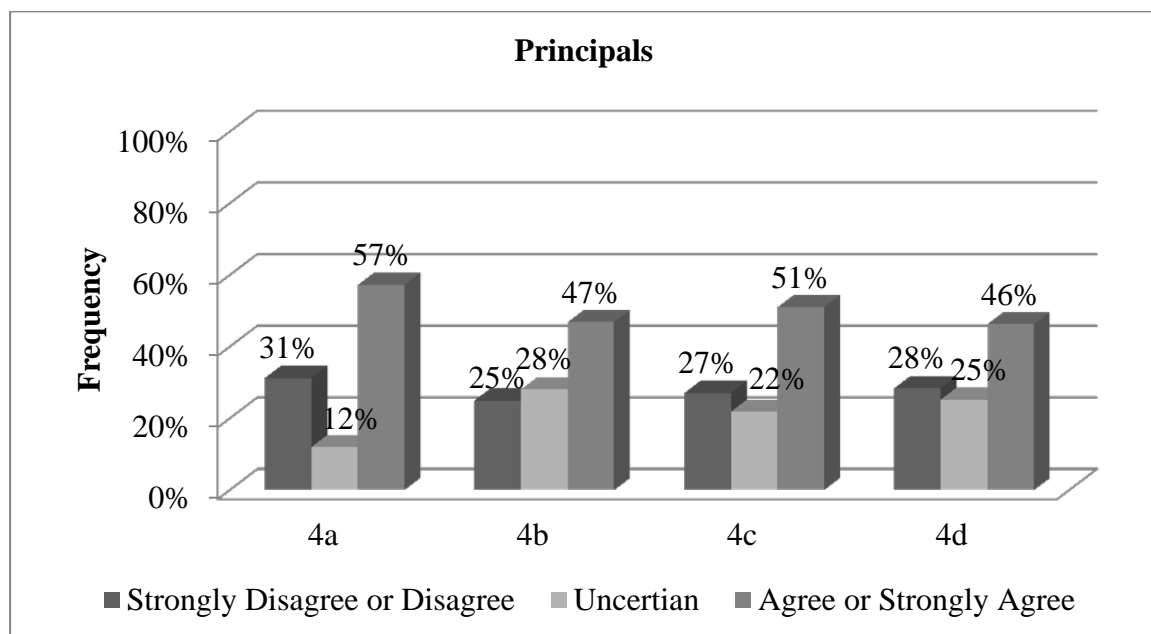


Figure 4

Principal Responses to Items 4a-4d



A large majority of the superintendents, 83%, agreed or strongly agreed with the statement that principal evaluation reflects what principals do on a daily basis, while only 57% of the principals agreed or strongly agreed. Twelve percent of principals who responded to the questionnaire were uncertain as to whether their evaluation reflected what they did on a daily basis.

When asked if superintendents and principals felt the criteria used to evaluate principals were researched-based, only 47% of principals indicated that they felt they were. However, 72% of the superintendents believed that their evaluation systems were researched-based. Although, there is a clear difference in perceptions on item 4b, a significant portion of each group, 24% of the superintendents and 28% of the principals were uncertain.

When asked in item 4c if principal evaluation improved job performance, 79% of superintendents and 51% of the principals believed that evaluation improved performance. However, a significant number, 27% of the principals disagreed with that statement. As with the other items, there was a large number, 17% of superintendents and 22% of principals that were uncertain.

Sixty-nine percent of the superintendents and 46% of the principals agreed or strongly agreed with the statement that principal evaluations used were effective. However, 28% of the principals disagreed or strongly disagreed with that statement. Like the other items in this section, there were nearly a quarter of the superintendents, 24%, and principals 25% that were uncertain.

Items 5a-5e in this section focused specifically on the perceptions of superintendents and principals related to the purpose of principal evaluation. Figure 5

and 6 represent responses of the superintendents and principals respectively. As in item 4 there tended to be more affirmative responses by superintendents and more diverse responses from the principals.

Figure 5

Superintendent Responses to Items 5a-5e

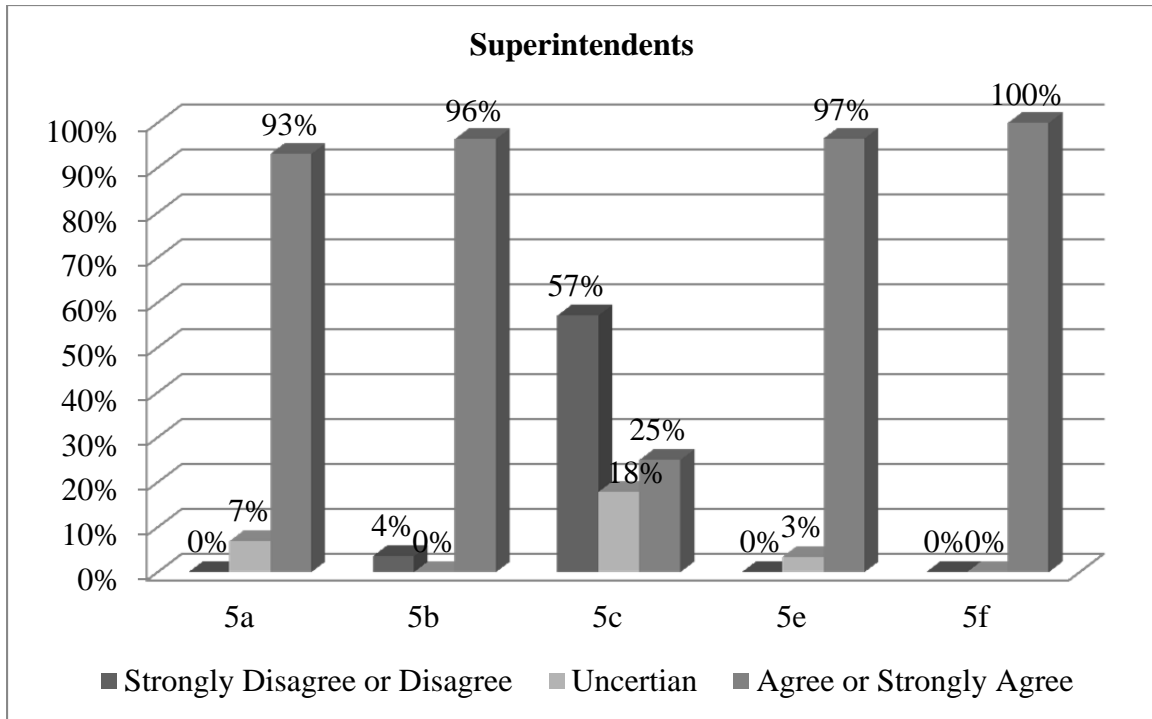
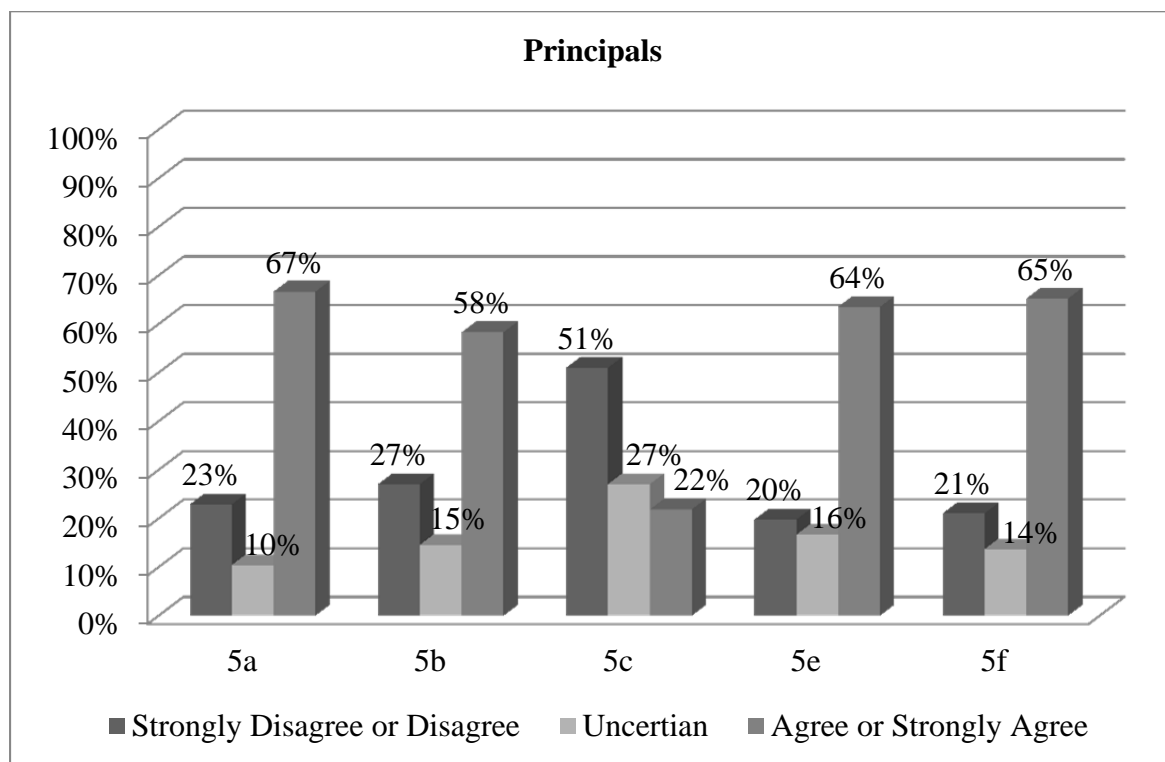


Figure 6

Principal Responses to Items 5a-5e

Item 5a asked if respondents agreed or disagreed with the statement that principal evaluation was focused on professional growth. Ninety-three percent of the superintendents and 66% of the principals felt that the principal evaluation was focused on professional growth. Likewise, when asked on item 5b if principal evaluation was used to assess job performance, a clear majority of superintendents, 93%, agreed or strongly agreed with that statement. However, only 58% of principals indicated that they agreed with that statement and 27% disagreed or strongly disagreed with this statement.

As indicated by the Mann-Whitney U, there was not a statistical difference in the responses between the principals and superintendents on item 5c which dealt with using principal evaluation to determine promotion or tenure. Both groups either strongly

disagreed or disagreed with the statement, 55% of superintendents and 51% of the principals.

Item 5e and item 5f were centered on student achievement and instructional quality. All but one of the superintendents, 97%, agreed or strongly agreed with the statement that principal evaluation was used to improve student achievement. All of the superintendents agreed or strongly agreed with the statement that principal evaluation was used to improve instruction in a school. Principals had a similarly strong affirmative response to principal evaluation being used to improve student achievement, however, it was not as strong as the superintendents with only 64% of principals agreeing or strongly agreeing. Similarly, only 65% of the principals agreed or strongly agreed with the notion that principal evaluation was used to improve instruction in a school.

Criteria. In the first phase of the study a template analysis was conducted in conjunction with the literature review. The researcher combined the findings of the literature review and template analysis and identified eight criteria that were either cited in research or by districts as being necessary to evaluate principals (Table 8).

Table 8

Mann-Whitney U Results Superintendent/Principal Comparison – Criteria

Items	Mann-Whitney U	z	Asymp. Sig. (2-tailed)
6a. Instructional Leadership	750.5	-3.626	0.000
6b. School Climate	759.5	-3.256	0.001
6c. Human Resources Administration	852.5	-2.776	0.003
6d. Assessing Instructional Quality	818.0	-3.154	0.002
6e. Organizational Management	829.5	-3.083	0.003
6f. Communication and Community Relations	996.0	-1.973	0.057
6g. Professionalism	1091.0	-1.367	0.301
6h. Student Achievement	838.0	-2.908	0.004

Notes: If z +/- 1.96, it's statistically significant at $p = 0.05$.

There was a statistical difference between the superintendents and principals in all but one of the criteria. Figure 7 and 8 display the responses of the superintendents and principals respectively.

Figure 7

Superintendent Responses to Items 6a-6h

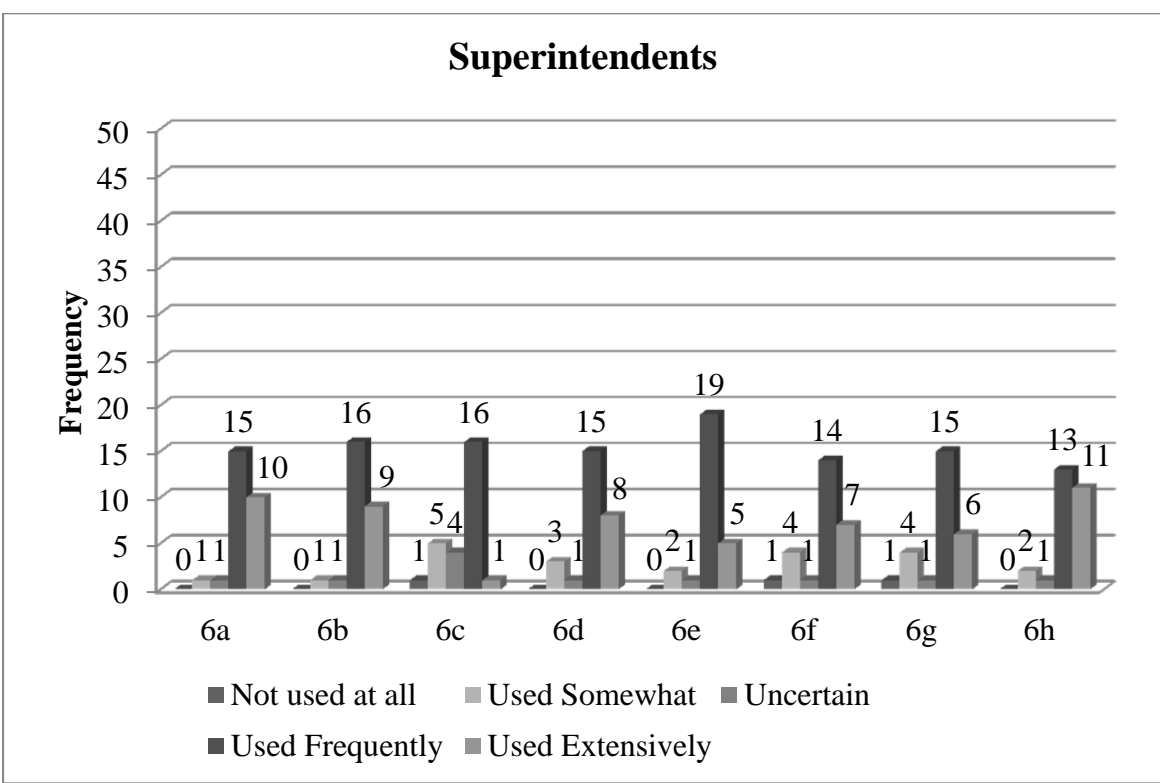
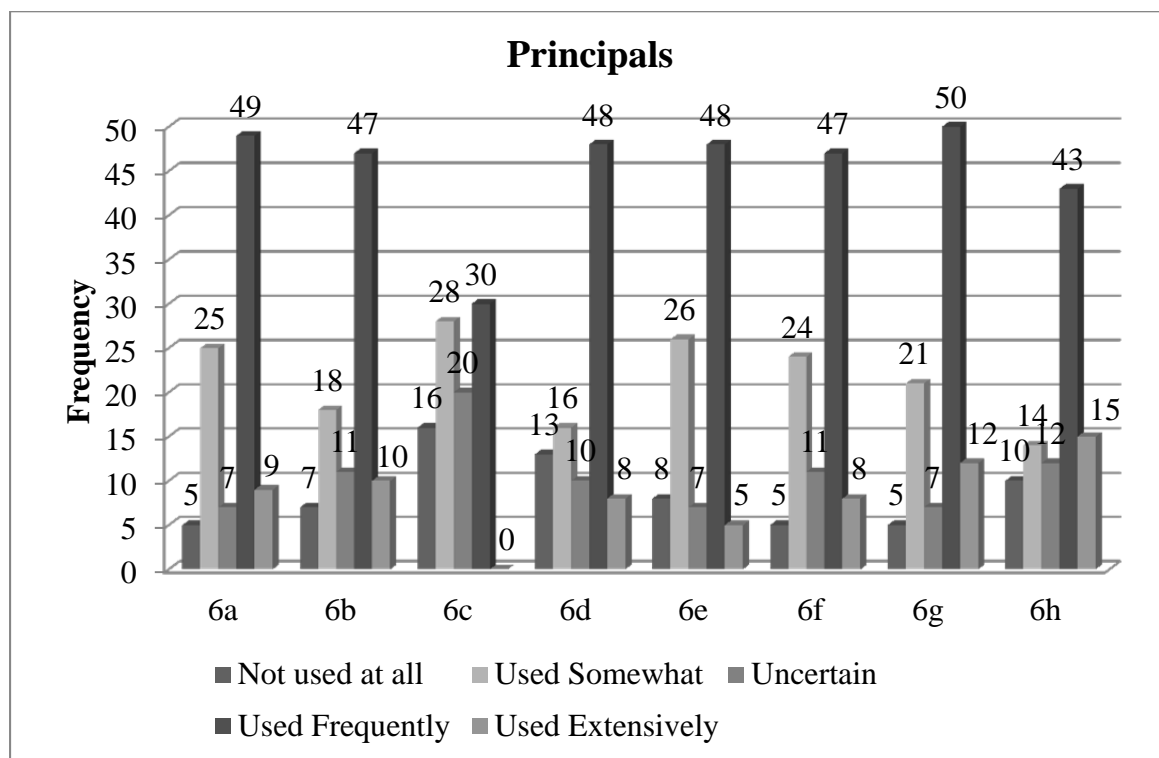


Figure 8

Principal Responses to Items 6a-6h

A majority of superintendents, 25 out of 27 or 93%, felt that instructional leadership was used frequently or extensively in a principal's evaluation. They held the same view in the area of school climate with 93% of superintendents responding that school climate was used frequently or extensively. A clear majority of principals felt that instructional leadership and school climate, 61%, were used frequently or extensively in their evaluations.

Both groups indicated that human resources administration, item 6c, was not used as frequently in principal evaluation, with 63% of superintendents and 32% of principal stating that it was used frequently or extensively. However, there was still a significant difference between the two groups.

Superintendents also held a view that assessing instructional quality, with 85% of the superintendents reporting it, was used frequently or extensively in the evaluation process. Principals differed in their view, with only 59% of them reporting that they felt it was used frequently or extensively. A similar difference was found in the area of organizational management, item 6e, where 89% of principals and 56% of principals felt that it was used frequently or extensively as a criterion in the evaluation process.

Although superintendents indicated that communication and community relations and professionalism, items 6f and 6g respectively, were routinely used in principal evaluation with 78% of the superintendents indicating they were, only 58% and 65% of principals reflected that same opinion. However, there was less of a statistical difference for item 6f and no statistical difference between the two groups related to professionalism.

The final criterion was student achievement. Sixty-two percent of principals felt that it was used frequently or extensively in their evaluations, whereas 89% of the superintendents indicated at that same frequency.

Artifacts and sources. The final quantitative section of the questionnaire dealt with what data is used to make evaluative judgments. As was cited in the template analysis and literature review, evaluations should be based upon multiple sources of data. Indeed, nearly every policy reviewed as part of the template analysis indicated that multiple sources of input or data were to be used to evaluate a principal. The questionnaire asked principals and superintendents to indicate what data principal evaluations were derived from. The questionnaire included 15 items grounded in the

template analysis or literature review and an option for respondents to add other sources or artifices that are used in the evaluation process (Table 9).

Table 9

Mann-Whitney U Results Superintendent/Principal Comparison – Artifacts and Sources

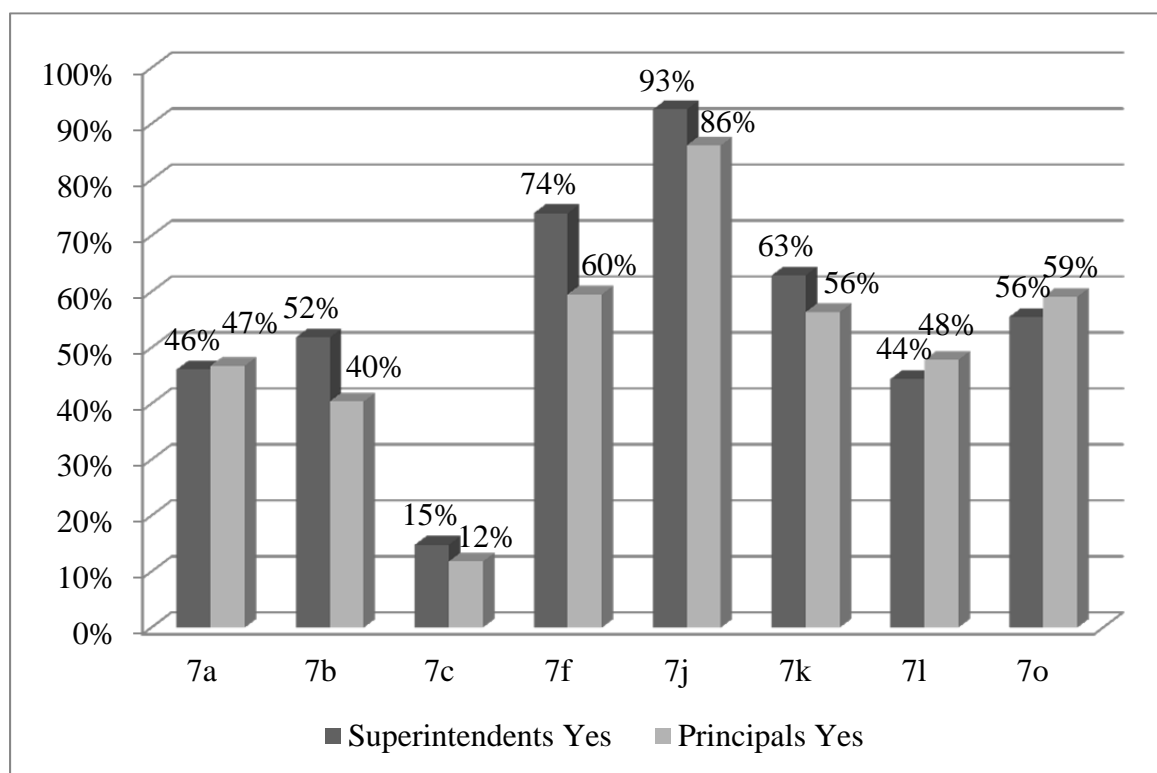
Items	Mann-Whitney U	z	Asymp. Sig. (2- tailed)
7a. Staff surveys	1258.5	-0.171	0.680
7b. Parent surveys	1130.5	-1.094	0.393
7c. Student surveys	1248.0	-0.229	0.881
7d. Student achievement on state assessments	1047.5	-2.586	0.098
7e. Student achievement on local assessments	909.5	-2.586	0.004
7f. School improvement plans	1163.0	-0.882	0.237
7g. Financial management	939.5	-2.500	0.018
7h. Self-reflection	911.5	-3.003	0.004
7i. Organizational skills	938.5	-2.844	0.006
7j. Perceptions of the supervisor	1202	-0.873	0.422
7k. Federal compliance	1204.0	-0.567	0.698
7l. Previous evaluations	1231.5	-0.634	0.549
7m. On-site observations	903.0	-2.646	0.013
7n. Portfolio of performance	1028.5	-2.076	0.090
7o. Community perceptions	1218.0	-0.372	0.574
8. How often is data collected through direct on-site observation by the person responsible for conducting principal evaluations?	561.5	-4.687	0.000

Notes: If z +/- 1.96, it's statistically significant at $p = 0.05$.

Unlike other sections of the questionnaire, there were more areas where there were no significant differences between groups. Out of the 15 items listed, principals and superintendents held similar views on eight of the items either being used or not being used as a source of information and data for principal evaluation (Figure 9). The strongest agreement was in the use of perceptions of the supervisor as part of the evaluation with 93% of superintendents and 86% of principals indicating perceptions were used in the evaluation process.

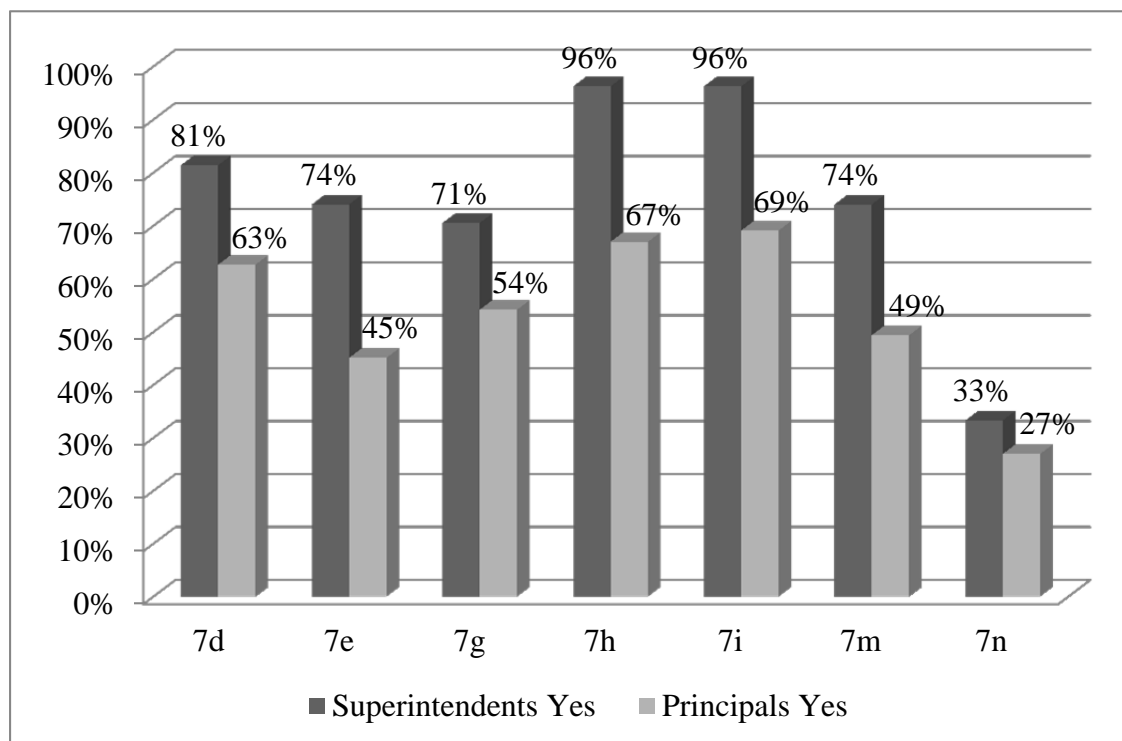
Figure 9

Comparison of Agreement on Eight Items



There were seven items that showed there was a statistical difference between the two groups (Figure 10). Those areas included the use of state assessment data, local assessment data, financial management, self-reflection, organizational skills, on-site observations and portfolios.

Figure 10

Comparison of Seven Items which were Statistically Different

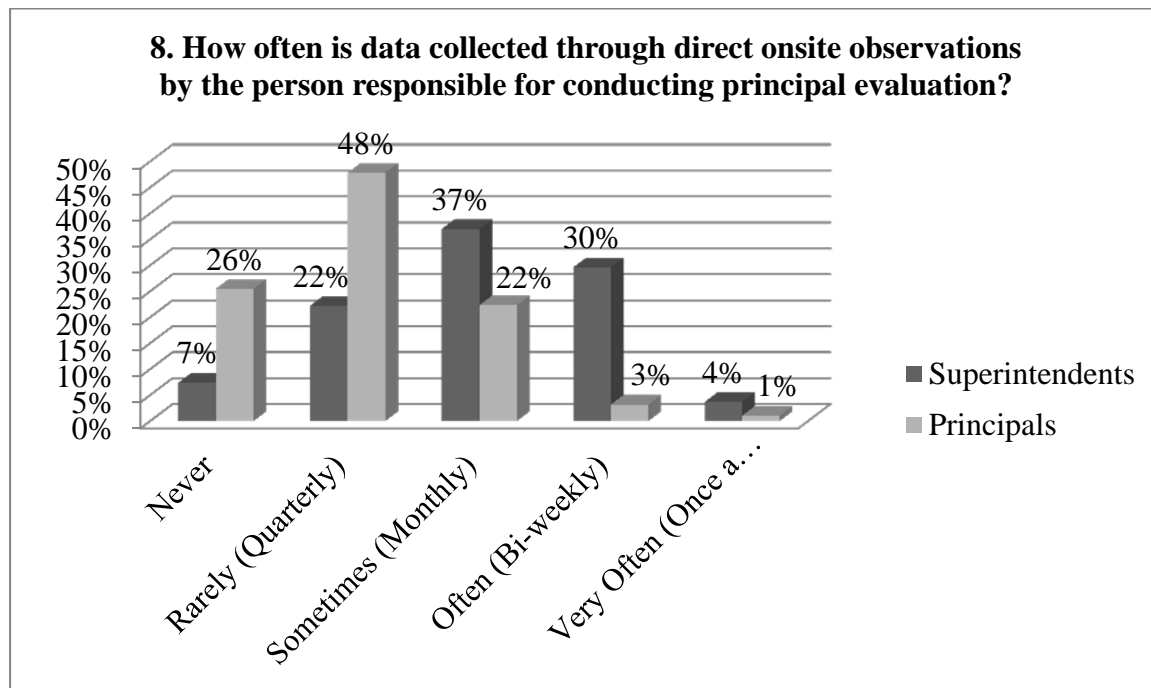
The largest difference was related to whether student achievement on local assessments was used as a data source for evaluation. Seventy-four percent of superintendents indicated that they used local assessments as a data sources while only 45% of principals reported that they were used. Another large difference was found in the area of principal's self-reflection being used in the evaluation process. Nearly all superintendents, 96% indicated that self-reflection was used in the process, where only 67% of principals felt that self-reflection was included. A similar difference was found in the use of organizational skills as part of the evaluation, with 96% of superintendents report they are used and only 69% of principals indicating their use.

Onsite observation was also an area of difference with 74% of superintendents reporting that this was a source of data for the evaluation and only 49% of principals

making a similar assertion. Figure 11 provides how frequently data is collected from onsite observations.

Figure 11

Comparison of Item 8



Seventy-four percent of principals indicate that they have either never had or rarely have an onsite observation by their supervisor compared to 29% of superintendents who indicate that onsite observations were rarely or not used in the evaluation process.

Seventy-one percent of the superintendents who responded to the survey indicated that they collect data at a minimum of monthly for the purpose of evaluation. Principals differ in that view where only 26% reported that they have the frequency of observation by their supervisor.

Participants were given an option to provide other sources of data used for principal evaluation. A number of items were given including e-mails from and to principals by stakeholders, public comment without verification or supporting

documentation, board trustee opinion, personal interviews, monthly school progress reports, Danielson's Framework, how well district initiatives are implemented, and how good a trainer the principal is.

Policy Development. Another important aspect of this study was to determine what role principals play in developing policies and practices related to principal evaluation. Participants were given four statements and asked to respond relative to their agreement or disagreement with the statement.

Table 9

Mann-Whitney U Results Superintendent/Principal Comparison – Principal Role in Evaluation

Items	Mann-Whitney U	z	Asymp. Sig. (2-Tailed)
9a. Principals participate in developing the district's principal evaluation policy.	688.5	-3.803	0.000
9b. Principals participate in selecting who will evaluate a principal's performance.	1028.5	-1.691	0.172
9c. Principals participate in determining what sources of data will be used in the evaluation.	945.0	-2.164	0.047
9d. Principals participate in developing professional development plans to improve performance.	825.5	-3.017	0.003

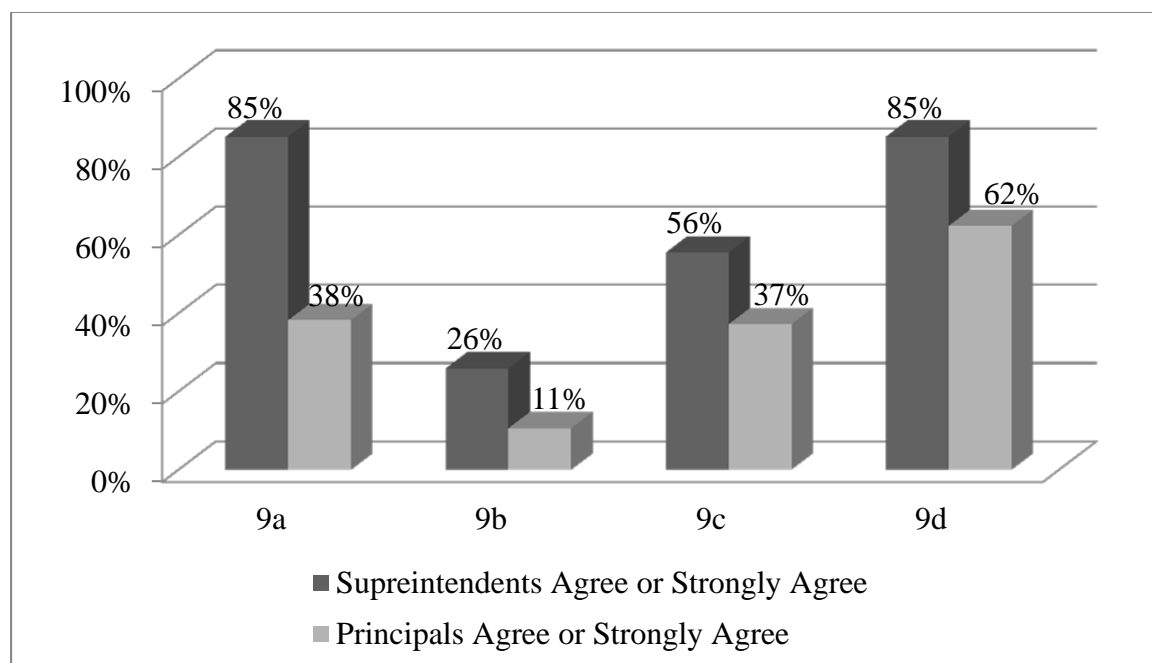
Notes: If z +/- 1.96 it's statistically significant at $p = 0.05$.

Three of the statements showed that there was a significant difference between the responses of principals and superintendents (Table 9). Principals and superintendents held the same perception that principals do not select their evaluator. However, there seemed to be a large disparity between principals and superintendents in whether principals participate in developing the district's principal evaluation policy. Eighty-five

percent of the superintendents agreed or strongly agreed with the statement that principal play a role in the development of the policy. However, only 38% of the principals indicated that they agreed or strongly agreed with that assertion (Figure 12). A smaller difference was found in the perception about principals selecting sources of data for their evaluation. Thirty-seven percent of the principals and 56% of superintendents felt that they had a role in selecting the data that was used in measuring their performance. A majority of principals 62% and superintendent 85% felt that principals participated in the process of developing professional development to improve their performance.

Figure 12

Comparison of Superintendents and Principals on Items 9a-9d



Qualitative Results. The final aspect of the questionnaire was open ended questions referring to the evaluation process. The first question asked participants to describe how principal evaluation worked in their district. In looking at the superintendent responses, it becomes very clear that they have a clear understanding of

the evaluation process in their district; their responses were detailed and lengthy. The principals were also detailed, however, there was more uncertainty regarding the process. For example, one respondent said, “I am not sure. No one has talked with me about evaluation. I think the superintendent will do one, but I have no idea when or what it will look like.” Another respondent stated, “I am not sure what the process is.” Although there were several that were not informed, most of the principals and all of the superintendents did outline their district’s practice. In analyzing the responses several key themes surfaced.

In reviewing the evaluation policies of the districts in the study area, only three district outlined professional inquiry as part of their policy (Table 1), further only one district policy specifically designated the use of growth plans as part of the evaluation process (Table 4). However, 15 superintendents and 30 principals described their evaluation process centering around a professional growth plan that incorporated goals. From the descriptions provided by the superintendents and principals these professional growth plans were the center of the entire evaluation process.

Another theme, which was associated with the growth plan, was self-reflection. In large part, 52 of the 72 principals who provided a written description of the evaluation process indicated that self-assessment and reflection were part of the evaluation process. They indicated that self-assessment and reflection was conducted either verbally or in writing and was used by the superintendent to develop the written final evaluation. Based upon the written comments of principals, self-assessment and reflection tended to be one of them most valued parts of the evaluation process. The following response given by one of the principals typified the responses of the principals. “The self-evaluation portion of

the process is probably the most valuable piece of the entire process.” Another respondent described the self-assessment process in these words. “I am asked to self-evaluate myself, and then sit in front of the superintendent, assistant superintendent, and all the district directors where I am asked questions, praised, scolded, and then given a written document.”

In all but two of the responses provided by principals in this study, the evaluation process culminated in a collaborative planning session or interview where performance, professional goals, school goals, and student achievement were discussed and a written formal evaluation document was generated and signed.

There appeared to be a sense that the evaluation process was a “hoop to be jumped through”. This was echoed by both superintendents and principals. Several respondents who were principals commented on the lack of evidence or sources for the evaluation. One reported that “I know he gets his information somewhere, but I don’t know where.”

What is very apparent is that both superintendents and principals seem to have the same perception about how evaluation occurs in their district. The following is a summation of their comments. The evaluation process starts off with a brief collaborative meeting towards the beginning of the year where the superintendent and principal discuss goals, usually focused on student achievement or school improvement. The principal works on those goals for most of the year and self-evaluates him or herself towards the end of the year. That self-evaluation is given to the superintendent who adds his or her comments. Another brief meeting is held where the principal is given the final evaluation.

Question 11 dealt with resources and professional development opportunities provided to principals to improve their performance. Twenty-two principals out of 75 and 15 superintendents out of 24 reported that professional development was provided to principals through administration meetings, and professional associations. Four principals out of the 75 who responded reported that their superintendent worked with them in the capacity of a mentor or professional coach.

On question 12, 33% of superintendents and 47% principals thought that the current principal evaluation used did not reflect what principals did on a daily basis. Principals indicated that while the criteria and instruments used to evaluate principals often times reflected the overarching vision of the position; they did not adequately reflect what principals do on a daily basis. One of the most frequently stated issues was the lack of onsite observations by the administrators who evaluate. Even in cases where principals indicated that there were frequent observations done at the school, those frequent observations did not adequately provide enough data to actually reflect the current job duties of a principal. Several of the respondents stated that the evaluation process did not reflect current practice due to being too narrowly focused on specific things. However, other respondents mentioned that the evaluation process was too general to adequately measure the daily responsibilities of a principal. Superintendents were more optimistic that the current evaluation process did reflect what happens on a daily basis; however, a number of them agreed with the principals regarding the evaluation process being outdated or too narrow or general.

Participants were given the opportunity to provide any other insights or information regarding principal evaluation in the final open-ended question of the questionnaire. Several unique things surfaced from those comments.

Principals felt that it was important that the policies and practices used to evaluate performance be developed by principals, superintendents, and school boards. Many of those who commented felt that principals should be evaluated by someone who had previous experience as a principal. One strong theme that emerged from the comments of principals was the need for evaluators, superintendents or other central office personnel, to spend more time observing a principal work in a school setting. Another strong theme was basing the evaluation in actual job performance and not in one-time events. A number of individuals who responded to the final question that were identified as principals indicated that far too often evaluative decisions or perceptions are formed after a principal makes a presentation to the board or a sports team wins a state competition. One of the respondents summarizes the perceptions of principals about evaluation very concisely when he states, "Being a principal is a tough job, don't make it tougher by overthinking it. If you want schools to be safe, then evaluations should be focused on climate and safety. If you want schools to be academically successful then evaluation should focus on academics. Don't make it so complex it is pointless."

Superintendents also provided additional comments, although far fewer than the principals. Of the twelve responses of superintendents, nearly all of them reflected the important role that principals play in helping schools succeed. Several indicated that principals are the critical component in helping increase student achievement at their

schools. A majority of those that provided a comment indicated that it was essential that the evaluation process be collaborative and include honest self-reflection.

Chapter V

Conclusion

Introduction

This study focused on four main research questions. This chapter contains a summation and findings for each of these questions.

1. What are the current practices in evaluating principals in school districts in southeastern Idaho and do they reflect the daily routines of a principal?
2. What criteria, artifacts and sources are being used to measure principal performance and how do they compare to best practices outlined in research?
3. What role does a principal play in principal evaluation?
4. How are the perceptions of principals and superintendents regarding principal evaluation in southeastern Idaho similar and different?

Summary of Results

Unlike the findings of Davis and Hensley (1999), this study also found that there is general consistency in the process used to evaluating principals. A majority of written policies currently used in the study area are similar, if not identical. The criteria and sources of data used to evaluate principals generally tended to be from the same sources. Nearly all of the participants, 82% of principals and 87% of superintendents, in the study described the process used to evaluate principals in a similar way which was comparable to Danielson (2007). A description of principal evaluation could be best summarized based upon the responses of this study as the following. Principals and supervisors met

toward the beginning of each school year to outline professional growth goals collaboratively. Principals collect data and self-assess their performance which are then given to the evaluator who reviews the self-assessment and another collected data. A concluding interview is held in mid to late spring where the written evaluation by the superintendent and central office administrators is given to the principal.

Although this methodology appears to be grounded in the research of Danielson, it appears that there are several areas of current practice that do not align between written policy, current practice and perception. The first of these deals with practice and written policy. In reviewing the policies and procedures of the districts, only four districts reported self-assessment or reflection as a part of their procedure or practice for evaluation (Table 4). Yet, when principals and superintendents were asked to describe the process nearly all indicated that self-assessment and reflection was a “critical component” of the process. When asked if principal evaluation was focused on professional growth, a clear majority 66% of principals and 93% of superintendents agreed that it was (Figure 5 & 6). A similar finding was found when asked if self-reflection was used as a source of information in the evaluation process, 67% of principals and 96% of superintendents indicated that they were. Based upon this information, there appears to be a clear disconnect between written policy and what is actually happening in practice.

The second question focused on what criteria, sources and artifacts are currently being used in the evaluation process. Although there was strong evidence that districts are using criteria that is aligned to behaviors that have been proven in research to effect

school culture and student achievement, there is a difference in perceptions of principals and superintendents about what criteria and evidence is used in evaluation.

During the template analysis, a majority of the districts who provided examples of their procedures and forms had a strong set of criteria that was aligned to those characteristic set forth by Stronge, Richard, and Catano (2008). Indeed, a majority of both superintendents and principals agreed that their district used the criteria similar to those found in research to evaluate principal's job performance (Figure 5 & 6). However, there was a statistical difference between principals and superintendents in their perceptions of how extensively criteria were used. Superintendents tended to have a strong affirmation that the criteria was being used frequently in the evaluation process, while principals were less certain and tended to feel that it was used less frequently. When asked if they agreed with the statement that the criteria used were researched based, a majority of the principals 53% were uncertain or disagreed with that statement (Figure 4).

There appears to a lack of communication between principals and superintendents about what criteria is used to assess job performance. This study did not collect sufficient data to be able to explore this divergence; however this would prove to be an interesting extension for future researchers.

Moore (2009) indicated the importance of multi-rater feedback as an essential component in evaluating principal's job performance. He asserts that using multiple sources of data and information to capture a clear picture of performance has historically been used in business, but has not been used in the educational sphere. This study found that most districts in the study area do use multiple sources of data to assess the job

performance of a principal. Principals and superintends both agreed that staff surveys, parent input, school improvement efforts, supervisor perceptions, previous evaluations, and community perceptions were all used in assessing performance. This is encouraging in large part, due to the fact that historically evaluation has often been based upon limited or no evidence other than evaluator perception. However, in looking at all of those areas where there was agreement, most of them are based upon external perceptions and little or no solid evidence. Although they give valuable feedback, staff and parent surveys are based upon individual's beliefs about the principal, and out of context, can paint an inaccurate picture of performance. Likewise, supervisor and community perceptions may be based on a handful of positive events and not reflect a holistic view of the performance. That is why it becomes critical to rely upon sources of data that are grounded in evidence to make evaluative judgments (Danielson, 2007).

Those areas that were evidenced based, such as, student achievement, federal compliance, on-site observations, portfolio of performance, and organizational skills were found to be areas of disagreement between the principals and superintendents. While superintendents indicated that these areas were used as sources of data, principals were less inclined to hold that same opinion (Figure 10). A clear example of this disagreement is in the use of on-site observations. Superintendents, 74%, indicate that they use on-site observation to help them formulate their evaluation decision. However, only 49% of principals feel that on-site observations are used as a source. Further, 71% of principals report that superintendents rarely if ever collect data from an on-site observation as part of their evaluation.

School districts in southeastern Idaho use criteria that appears to be aligned to research and judgments made about performance related to that criteria is based on multiple sources of data. However, those sources are oftentimes based upon limited evidence or perceptual data.

The next question dealt with the role the principal plays in principal evaluation. The researcher explored what role the principal plays in the actual development of the policies and procedures used to evaluate their performance and also the role they play in the actual evaluation process.

The role of the principal in developing the evaluation policies and practices was very limited as reported by both the principals and superintends. Both groups reported that principals do not select who is involved in the evaluation process and principals do not select the sources used to evaluate performance. Similar to the other areas of the study, there was a clear divergence between the principals and superintendents with regards to the role the principal played in writing the evaluation policy. Superintendents felt that principals do participate in developing the policy for evaluation; however, principals did not hold that same view (Figure 12).

The final question focused on determining if there were differences between principal and superintendent perception about evaluation. In almost every aspect of the questionnaire there was a statistical difference between the two groups. Superintendents held a more positive view towards the evaluation process. As a group, superintendents responded more frequently in the affirmative than principals did. Principals tended to be more uncertain about different aspects of the questionnaire. The uncertainty seemed to also be reflected in their comments to the open-ended questions. Principals indicated in

their responses that it was essential that they participate in the evaluation process and that those evaluating principal job performance have actually been principals before.

Conclusions

The study found that school districts in southeastern Idaho have a strong sense of developing evaluation systems that are focused on professional growth and self-assessment and reflection. Although many of the policies and practices current adopted by districts are not aligned to a districts' current practices, much of what districts are doing is aligned to best practices in research regarding evaluation.

There is still a strong difference of perception between principals and superintendents about the purposes and criteria used to assess job performance. This is largely due to the lack of participation by principals in establishing the evaluation policy and practice. There is a need for districts to revisit their current policies and collaboratively develop new policies with principals. Through this process a clearly defined set of criteria and sources of evidence should be developed to help principals more fully participate in the evaluation experience.

Recommendations for Further Research

This study focused on the current practices and perceptions of superintendents and principals, however, further study is needed to continue to enhance and advance the dialogue. A critical area that needs further exploration is the concept brought forth by Reeves (2004) regarding the "knowing doing gap". One of the biggest discrepancies found in this study was the difference between what was in written policy and what was being reported by superintendents and principals in the field. Further, in almost every area there was a statistical difference between principals and superintendents about

almost every aspect of the evaluation process. These discrepancies allude to the possibility that what principals and superintendents believe is happening with regard to principal evaluation is actually not happening at all.

Further qualitative studies are needed to help explore how principal evaluation looks from both an evaluator's perspective and the perspective of those that are being evaluated. This would help advance the conversation on evaluation and help triangulate actual practice comparatively to written expectations.

Another potential area for further study is in whether principal evaluation actually contributes to improved student achievement. As was mentioned in the literature review, principals have been found to at least have an indirect impact on student achievement, but further investigation as to whether the evaluation process contributes to improved student achievement would be very valuable to the dialogue about job performance. While there is limited evidence to suggest that evaluation can change behavior, it is far from definitive. Surprisingly, there is very little quantitative evidence to suggest that job assessment or evaluation have any impact on student performance.

In the open-ended portion of the questionnaire, the notion of professional development taking place at monthly or bi-monthly administration meetings was asserted by both the superintendents and principals. Further research is needed on the effectiveness of principal professional development, specifically, if the current methods of principal professional development are contributing to improved job performance and student achievement performance.

Finally, additional study is needed on principal mentoring and professional development. In this study, three districts reported that they provided ongoing

professional development for principals. In investigating those programs, they were mostly attending annual professional organizations conferences, book studies, or other non-coordinated professional development events. A vast majority of the principals in the study indicated that they had no organized focused professional development on improving their job performance. Studies are needed to broaden the discussion about mentoring and administrative coaching to help principals improve and impact student and teacher performance.

Implications for Professional Practice

Although additional research on the effect principal evaluation has on changing behavior to impact student achievement, there are several implications from this study that would benefit the current professional practice. The first is the need for greater collaboration between principals and superintendents in developing a clear and concise evaluation process that results in behavioral changes. Similar to what Thomas, Holdaway, and Ward (2000) found, principals in this study still have a limited role in developing the processes and tools used to evaluate them. Although principals reported participating in the evaluation process through self-assessment and reflection, the process is still an externally developed process that happens to the principal annually. This was an area that both principals and superintendents had a similar view, that principals currently had no role in developing the policies and practices for evaluation. As a result, principals differed in their view about the purpose of the evaluation system with superintendents.

This leads to another implication which is a need for stronger communication between principals and superintendents regarding the purpose and process of evaluation. Principals and superintendents should be able to articulate the criteria used to evaluate

their job performance and also cite the sources used to make judgments about performance.

Conclusion

Over the past century the role of education has become increasingly complex and overburdened with new mandates and requirements (Vollmer, 2010). A century ago, education was responsible for providing the basics of reading, writing, and arithmetic. Today, education has expanded to not only include the those three basic content areas, but also, music, social studies, science, bullying prevention, personal financial literacy, health and wellness, media literacy, technology literacy, and many others. As the burdens and demands placed upon education have increased, so has the job complexity of those who operate in education. The role of the principal has evolved from being largely managerial and disciplinary with relatively few responsibilities to one focused on instructional leadership and continuous improvement with hundreds of responsibilities. Indeed, so demanding is the role of the principalship, it caused DuFour, DuFour, and Eaker (2008) to write, “no single person could address all the responsibilities principals have been asked to shoulder” (p. 311).

This study sought to determine the current reality of principal evaluation in southeastern Idaho. What was found was a system that is still struggling to evolve to meet the demands of today’s educational expectations. There is a knowing doing gap in southeastern Idaho related to evaluation. Many policies and procedures are outdated, or if they are more current, they do not reflect the actual practices being implemented. Although, the principal evaluation processes studied offer an optimistic view that principals are an active part of the evaluation process, principals still play a very limited

role in setting policies that govern their evaluation and the sources of data used to assess their performance.

If school districts expect to impact student achievement using the evaluation process as a component for improvement, then superintendents and principals need to work collaboratively to redefine how principals are evaluated to better align with the current demands of the educational system. School districts need to develop criteria that define the behaviors exhibited by principals that have been found to improve student achievement. These criteria need to be jointly defined by principals and superintendents and clearly communicated to all administrators in a district. Not only do the criteria need to be clearly defined in policy, but they need to be reflected in practice, thus averting the knowing-doing gap. The evaluation systems developed by principals and superintendents need also to draw from multiple sources of data and artifacts of practice. Evaluation should be based on a collection on facts rooted in evidence and data, rather than on opinion and assumption.

As was stated at the beginning of this study by Jim Collins (2001), “You must ...have the discipline to confront the most brutal facts of your current reality, whatever they might be” (p. 13). The time is now for school districts to look at their current reality and face the brutal facts. It is time to begin the process to collaboratively set a vision for what principal evaluation should be. The purposes of principal evaluation need to continue to develop and grow in the dialogue of educational research, so that a better understanding and system of evaluation can emerge.

References

- Andrews, C. (1990). Evaluating principals. *Research Roundup*, 6(2), 3-6. (ED318131).
- Archer, J. (2002). Principals: So much to do, so little time. *Education Week*, 21(31), 1-20.
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology*, 49(1), 141–168
- Babo, G. (2009). The ISLLC standards' "footprints" and principal evaluation: A national study of school superintendents. *Connexions*. Retrieved from <http://cnx.org/content/m33273/1.3/>.
- Baruch, Y. (1999). Response rate in academic studies-A comparative analysis. *Human Relations* 52(4), 412-438. doi: 10.1177/001872679905200401
- Blum, R. E., Butler, J. A., & Olsen, N. L. (1987). Leadership for excellence: Research-based training for principals. *Educational Leadership*, 25-29.
- Brady, L. (2012). *Principal evaluation: A description of current practices* (Doctoral dissertation). Retrieved from http://sdsu-dspace.calstate.edu/bitstream/handle/10211.10/2031/Brady_Leighangela.pdf?sequence=1
- Brooks, M., & Voss, J. (2008). *An analysis of the effectiveness of the 2003 Missouri Department of Elementary and Secondary Education Performance-Based Principal Evaluation as compared to McRel Balanced Leadership Framework*

- (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI 3324142)
- Brown, L. (2002, March). *Issues in school leadership and school effectiveness: A focus on Trinidad and Tobago*. Paper presented at the International Conference on Problems and Prospective of Education in Developing Countries, Bridgetown, Barbados, West Indies.
- Bulach, C., Boothe, D., & Pickett, W. (2006). Analyzing the leadership behavior of school principals. Retrieved from Connexions Web site:
<http://cnx.org/content/m13813/latest/>
- Bush, T., & Chew, J. (1999). Developing human capital: Training and mentoring for principals. *Compare*, 29(1), 42-52.
- Catano, N., & Stronge, J. H. (2007). What do we expect of school principals? Congruence between principal evaluation and performance standards. *International Journal of Leadership in Education*, 10(4), 379-399.
- Collins, J. (2001). *Good to great*. New York, NY: HarperCollins Publishers Inc.
- Conca, J. (2008). *Principal evaluation: A qualitative study of public school superintendents' perceptions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3347467)
- Condon, C., & Clifford, M. (2010). *Measuring principal performance: How rigorous are commonly used principal performance assessment instruments?* [White paper]. Retrieved from Quality School Leadership website:
<http://www.learningpt.org/pdfs/QSLBrief2.pdf>

- Cotton, K. (2003). *Principals and student achievement: What the research says*. Alexandria, VA: ASCD.
- Council of Chief State School Officers. (1996). *Interstate school leaders licensure consortium: Standards for school leaders*. Washington, DC: Author.
- Creswell, J. W. (2008). *Instructional research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: ASCD.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD.
- Davis, S. H., & Hensley, P. A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education*, 13(4), 383-403.
- Davis, S., Kearney, K., Sanders, N., Thomas, C., & Leon, R. (2011). *The policies and practices of principal evaluation: A review of the literature*. San Francisco, CA: WestEd.
- Derrington, M., & Sharratt, G. (2008). Evaluation of school principals using interstate school leaders licensure consortium (ISLLC) standards. *AASA Journal of Scholarship & Practice*, 5(3), 20-29.
- DuFour, R., DuFour, R., & Eaker, R. (2008). *Revisiting professional learning communities at work: New insights for improving schools*. Bloomington, IN: Solution Tree Press.

- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, IN: National Educational Service.
- DuFour, R. & Marzano, R. (2011). *Leaders of learning: How district, school, and classroom leaders improve student achievement*. Bloomington, IN: Solution Tree Press.
- Eaker, R., DuFour, R., & DuFour, R. (2002). *Getting started: Reculturing schools to become professional learning communities*. Bloomington, IN: National Educational Service.
- Ediger, M. (1998). *Appraising the school principal*. Retrieved from ERIC database. (ED427023)
- Feldman, J., & Ouimette, M. (2004). *Examining the turning points comprehensive middle school reform model: The role of local context and innovation*. Retrieved from ERIC database. (ED509779)
- Gaziel, H. (2003). Images of leadership and their effect upon school principals' performance. *International Review of Education*, 49(5), 475-486.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Ginsberg, R., & Berry, B. (1990). The folklore of principal evaluation. *Journal of Personnel Evaluation in Education*, 3(3), 205-230.
- Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., & Carson, B. (2009a). The evaluation of principals: What and how do states and urban districts assess leadership? *The Elementary School Journal*, 110(1), 19-39.

- Goldring, E., Porter, A., Murphey, J., Elliot, S., & Cravens, X. (2009b). Assessing learning-center leadership: Connections to research, standards, and practices. *Leadership and Policy in Schools*, 8(1), 1-36.
- Grissom, J., & Loeb, S. (2009). *Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills* [Working paper]. Retrieved from CALDER website: http://www.caldercenter.org/upload/Working-paper-35_FINAL-2.pdf
- Hall, P. (2008). Building bridges: Strengthening the principal induction process through intentional mentoring. *Phi Delta Kappan*, 89(6), 449-452.
- Harrison, W. C. (1988, April). *The status of evaluation of principals: Administrator perceptions and opportunities and obstacles to improving the process*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Harrison, W. C., & Peterson, K. D. (1987, April). *Complexities in the evaluation of principals: The relationship between satisfaction with evaluation processes, criteria, and sources of information*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.
- Hart, A. W. (1992, April). *The social and organizational influence of principals: Evaluating principals on organizational criteria*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Hatrick, E. B. (2008). A climate for success. *The School Administrator*, 9(65), 1-10.
- Heck, R. H., & Glasman, N. S. (1993). Merging evaluation and administration. *Peabody Journal of Education*, 68(2). 132-142.

- Hessel, K., & Holloway, J. (2002). *A framework for school leaders: Linking the ISLLC standards to practice*. Princeton, NJ: Educational Testing Service.
- Ibukun, W. O., Oyewole, B. K., & Abe, T. O. (2011). Personality characteristics and principal leadership effectiveness in Ekiti State, Nigeria. *International Journal of Leadership Studies*, 6(2), 247-262.
- Kafka, J. (2009). The principalship in historical perspective. *Peabody Journal of Education*, 84(1), 318-330.
- Kearney, K. (2005). Guiding improvements in principal performance. *Leadership*, 18-21.
- Kimball, S. M., Milanowski, A., & McKinney, S. A. (2009). Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial. *Leadership and Policy in Schools*, 8(1), 233-263.
doi:10.1080/15700760802416099
- Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oaks, CA: Corwin Press.
- Kuzma, D. J. (2004). *The characteristics of effective middle school principals as perceived by middle school teachers and the principals for whom they work* (Unpublished doctoral dissertation). Seton Hall University, South Orange, New Jersey. Retrieved from
[http://domapp01.shu.edu/depts/uc/apps/libraryrepository.nsf/resourceid/2994BB15FCD3F2B485256E6600711D92/\\$File/kuzma-deborah-j_Doctorate.pdf?Open](http://domapp01.shu.edu/depts/uc/apps/libraryrepository.nsf/resourceid/2994BB15FCD3F2B485256E6600711D92/$File/kuzma-deborah-j_Doctorate.pdf?Open)
- Lashway, L. (2003). *Improving principal evaluation*. Retrieved from ERIC database. (ED482347)

- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-386.
- Longenecker, C. O. & Nykodym, N. (1996). Public sector performance appraisal effectiveness: A case study. *Public Personnel Management, 25*(1), 151-164.
- Marcoux, J., Brown, G., Irby, B. J., & Lara-Alecio, R. (2003, April). *A case study on the use of portfolios in principal evaluation*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Marshall, C., & Rossman, G. B. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks, CA: SAGE Publication, Inc.
- Martineau, P. (2012). Principles of good principals: Effective leadership brings a board's vision to the school level. *Education Digest: Essential Readings Condensed for Quick Review 77*(8), 53-58.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: ASCD.
- Marzano, R. J., & Waters, T. (2009). *District leadership that works: Striking the right balance*. Bloomington, IN: Solution Tree Press.
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.
- McEwan, E. K. (2003). *Ten traits of highly effective principals: From good to great performance*. Thousand Oaks, CA: Corwin Press, Inc.
- McGahie, W. C. (1991). Professional competence evaluation. *Educational Researcher, 20*(1), 3-9.

- Mitchell, C., & Castle, J. B. (2005). The instructional role of elementary school principals. *Canadian Journal of Education*, 28(3), 409-579.
- Moore, B. (2009, January/February). Improving the evaluation and feedback process for principals. *Principal*. Retrieved from http://www.naesp.org/resources/2/Principal/2009/J-F_p38.pdf
- Murphy, J. A., & Pimental, S. (1996). Grading principals: Administrator evaluations come of age. *Phi Delta Kappan*, 78(1), 74-81.
- New Leaders for New Schools. (2010). *Evaluating principals: Balancing accountability with professional growth* [White paper]. Retrieved from Illinois State Board of Education: http://www.isbe.net/peac/pdf/eval_principals.pdf
- Nor, S., & Roslan, S. (2009). Turning around at-risk schools: What effective principals do. *International Journal on School Disaffection*, 6(2), 21-29.
- Peterson, K. D., & Deal, T. E. (1998). How leaders influence the culture of schools. *Educational Leadership*, 56(1), 28-30.
- Prestine, N. (2008, October). *Principal induction programs: What the teacher induction literature can (and cannot) tell us*. Paper presented at the University Council of Educational Administration annual convention, Orlando, FL.
- Raymer, A. N. (2006). *Principal leadership and school culture in public schools: Case studies of Piedmont North Carolina elementary schools* (Unpublished doctoral dissertation). University of North Carolina, Greensboro, North Carolina.
- Reeves, D. B. (2006). *The learning leader: How to focus school improvement for better results*. Alexandria, VA: ASCD.

- Reeves, D. B. (2004). *Assessing educational leaders: Evaluating performance for improved individual and organizational results* (1st ed.). Thousand Oaks, CA: Corwin Press.
- Reeves, D. B. (2009). *Leading change in your school: How to conquer myths, build commitment, and get results*. Alexandria, VA: ASCD.
- Robinson, V., Eddy, D., & Irving, E. (2006). Catering for diversity in a principal induction program. *School Leadership and Management*, 26(1), 149-167.
doi:10.1080/13634230600589691
- Rooney, J. (2005). School culture: An invisible essential. *Educational Leadership*, 61(4), 86.
- Schrader, B. W., & Steiner, D. D. (1996). Common comparison standards: An approach to improving agreement between self and supervisory performance ratings. *Journal of Applied Psychology*, 81(6), 813-820.
- Schulte, D. P., Slate, J. R., & Onwuegbuzie, A. J. (2010). Characteristics of effective school principals: A mixed-research study. *The Alberta Journal of Educational Research*, 56(2), 172-195.
- Shannon, G. S., & Bylsma, P. (2007). *The nine characteristics of high-performing schools: A research-based resource for schools and districts to assist with improving student learning* (2nd ed.). Olympia, WA: Office of the Superintendent of Public Instruction.
- Sterrett, W. (2011). *Insights into action: Successful school leaders share what works*. Alexandria, VA: ASCD.

- Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.
- Stine, D. O. (2001, April). *Developing an evaluation system to improve principal performance and accountability*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Studebaker, G.L. (2000) *A comparative study of Indiana superintendent's perceptions of principal evaluation instruments* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9965722)
- Stufflebeam, D., & Nevo, D. (1993). Principal evaluation: New directions for improvement. *Peabody Journal of Education*, 14(3), 215-240.
- Stronge, J. H., Richard, H. B., & Catano, N. (2008). *Qualities of effective principals*. Alexandria, VA: ASCD.
- Sun, M., & Youngs, P. (2009). How does district principal evaluation affect learning-centered principal leadership? Evidence from Michigan school districts. *Leadership and Policy in Schools*, 8, 411-445. doi:10.1080/15700760902890490
- Tanner, D. (2012). *Using statistics to make educational decisions*. Thousand Oaks, CA: SAGE Publication, Inc.
- Taylor, L. M. (2011). *The importance of school culture in the advancement of 21st century learning environments: A qualitative study* (Unpublished doctoral dissertation). West Virginia University, Morgantown, West Virginia.
- The Education Alliance at Brown University. (2004). *Making the case for principal mentoring* [White paper]. Retrieved from The Education Alliance at Brown University: <http://www.alliance.brown.edu/pubs/pln/prncpalmntrg.pdf>

- Thomas, D. W., Holdaway, E. A., & Ward, K. L. (2000). Policies and practices involved in the evaluation of school principals. *Educational Assessment, Evaluation and Accountability, 14*(3), 215-240.
- Toler, B. B. (2006). *A comparative study of principal performance evaluation in the commonwealth of Virginia* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3258819)
- Vollmer, J. R. (2010). *Schools cannot do it alone*. Fairfield, IA: Enlightenment Press.
- Walker, A., & Qian, H. (2006). Beginning principals: Balancing at the top of the greasy pole. *Journal of Educational Administration, 44*(4), 297-309.
doi:10.1108/09578230610674921
- Whitaker, T. (2003). *What great principals do differently: Fifteen things that matter most*. Larchmont, NY: Eye on Education.
- Williams, E., Persaud, G., & Turner, T. (2008). Planning for principal evaluation: effects on school climate and achievement. *Educational Planning, 17*(3), 1-11.

Appendix A

Template Analysis of Evaluation Policies

	Policies	Criteria	Sources of Data	Professional Inquiry	Self-assessment	Mentoring Collaboration	None
Districts 1	X	X	X				
District 2	X	X	X	X		X	
District 3	X	X	X				
District 4	X	X	X				
District 5	X	X					
District 6	X	X					
District 7	X	X					
District 8	X	X					
District 9	X	X					
District 10	X						
District 11	X						
District 12	X	X	X				
District 13	X						
District 14	X						
District 15	X						
District 16	X		X				
District 17	X	X	X				
District 18	X	X	X	X	X	X	

District 19	X	X	X		X		
District 20	X	X					
District 21	X	X	X		X		
District 22	X	X	X		X		
District 23	X	X					
District 24	X	X					
District 25	X	X		X		X	
District 26	X	X	X				
District 27	X	X	X				
District 28	X	X	X				
District 29	X	X	X				
District 30	X	X					
District 31							X
District 32							X
Total	30	24	15	3	4	3	2

Appendix B

Principal Evaluation Questionnaire

Informed Consent

I, Donald Bingham, am conducting a survey as part of my doctoral studies at Northwest Nazarene University focused on the topic of principal evaluation in southern Idaho. The primary purpose of this study is to identify current practices being used to evaluate principals in southern Idaho and how accurately principal evaluation reflect what principals do on a daily basis. This 10 minute survey is voluntary. You may choose not to answer any questions that you find embarrassing or offensive and you may discontinue taking the survey at any time. If you have any undue stress or anxiety as a result of taking this survey or other concerns about your rights as a participant, I am available by calling 208-745-6693 ext. 1102 or you may contact my chair, Dr. Paula Kellerer, at 208-467-8729.

Survey results from principals and superintendents are anonymous and not even the researcher will be able to connect Responses to individuals. Thank you.

By continuing in this survey, you give your consent to participate in this study.
(I give consent or I do not give my consent)

Demographics

1. What was your position prior to your current position as a principal?
(Superintendent, Assistant Superintendent or Director, Principal, Vice Principal, or Other)
2. How long have you been in your current position?
(1 year or less, 2-10 years, 11-20 years, 21-30 years, or 31+ years)
3. Who is responsible in your school system for evaluating principals?
(Superintendent, Assistant Superintendent, Other Central Office Administrator, Self-Evaluation, or Other)

Purposes and Practices

Response Choices for Purposes and Practices Questions

Strongly Disagree
Disagree
Uncertain
Agree
Strongly Agree

4. In my district:
 - a) Principal evaluation reflects what principals do on a daily basis.
 - b) The criteria used in principal evaluation are researched-based.

- c) Principal evaluation improves principal job performance.
 - d) Principal evaluation is effective.
5. The purpose of principal evaluation in my district is:
- a) To promote professional growth as a principal.
 - b) To assess job performance as measured by a set of clearly communicated criteria or standards.
 - c) To determine promotion or tenure.
 - d) To determine pay for performance awards for principals.
 - e) To improve student achievement.
 - f) To improve instruction in schools.

Criteria

Response Choices for Criteria Questions

- Not Used at All
- Used Somewhat
- Uncertain
- Used Frequently
- Used Extensively

6. To what extent are the following criteria used to evaluate principals:
- a) Instructional Leadership
 - b) School Climate
 - c) Human Resources Administration
 - d) Assessing Instructional Quality
 - e) Organizational Management
 - f) Communication and Community Relations
 - g) Professionalism
 - h) Student Achievement

Artifacts and Sources

Response Choices for Sources or Artifacts

- Yes
- No

7. The following sources or artifacts are used to evaluate a principals job performance:
- a) Staff surveys
 - b) Parent surveys
 - c) Student surveys
 - d) Student achievement on state assessments
 - e) Student achievement on local assessments
 - f) School improvement plans
 - g) Financial management
 - h) Self-reflection

- i) Organizational skills
 - j) Perceptions of the supervisor
 - k) Federal compliance
 - l) Previous evaluations
 - m) On-site observations
 - n) Portfolio of performance
 - o) Community perceptions
 - p) Honors and awards the school receives
 - q) Other
8. How often is data collected through direct on-site observations by the person responsible for conducting principal evaluation? (Never, Rarely (Quarterly), Sometimes (Monthly), Often (Bi-weekly), or Very Often (Once a week))

Policy Development

Response Choices for Policies

- Strongly Disagree
 - Disagree
 - Uncertain
 - Agree
 - Strongly Agree
9. Principals participate in:
- a) Developing the district's principal evaluation policies.
 - b) Selecting who will evaluate a principal's performance.
 - c) Determining what sources of data will be used in the evaluation.
 - d) Developing professional development plans to improve performance.

Qualitative Questions

10. In your own words, briefly describe the process used in your district to evaluate principals.
11. Please describe what resources and support your district currently provides principals to improve their performance as a school leader.
12. Do you feel that the evaluation process currently used in your district accurately represents what principals do on a daily basis? Please explain why or why not.
13. Please provide any other comments or thoughts that you consider to be relevant to the topic of evaluation of principals.

Appendix C

Forms of Evaluation Criteria

	Defined Performance Levels	Likert-Scale	Artifacts	Narrative	No Description Provided
Districts 1					X
Districts 2					X
Districts 3					X
Districts 4					X
Districts 5					X
Districts 6					X
Districts 7					X
Districts 8					X
Districts 9					X
Districts 12					X
Districts 17					X
Districts 18			X	X	
Districts 19		X		X	
Districts 20				X	
Districts 21		X		X	
Districts 22	X			X	
Districts 23	X			X	
Districts 24		X		X	
Districts 25					X
Districts 26		X		X	
Districts 27		X	X	X	

Districts 28	X			X	
Districts 29		X		X	
Districts 30		X		X	
Totals	3	7	2	12	12

Appendix D

Comparison of Research-based Criteria and District Policies

	Inst. Lead.	Climate	Staff Dev.	Staff Eval.	Manag.	Comm. Relations	Achieve.	None
Districts 1								X
Districts 2								X
Districts 3								X
Districts 4								X
Districts 5								X
Districts 6								X
Districts 7								X
Districts 8								X
Districts 9								X
Districts 12								X
Districts 17								X
Districts 18	X	X	X	X		X	X	
Districts 19	X	X	X	X		X	X	
Districts 20	X	X	X	X			X	
Districts 21	X	X				X	X	
Districts 22	X	X				X	X	
Districts 23	X	X	X	X	X	X	X	
Districts 24		X			X	X	X	
Districts 25								X
Districts 26	X	X	X	X	X		X	
Districts 27				X		X	X	

Districts 28	X	X	X	X	X	X	X	
Districts 29	X	X	X	X	X	X	X	
Districts 30								X
Totals	8	10	6	8	5	9	11	12

Appendix E

Disaggregated Multiple Sources of Data

	Student Achievement	Parent Input	Growth- Plans	Self- Assessment	Staff Perceptions	Evaluation Criteria
Districts 1	X	X				X
Districts 2		X				X
Districts 3	X	X				X
Districts 4		X				X
Districts 12		X				X
Districts 16	X	X				X
Districts 17	X	X				X
Districts 18		X	X			X
Districts 19		X		X	X	X
Districts 21	X	X		X		X
Districts 22	X	X		X		X
Districts 26	X	X				X
Districts 27		X			X	X
Districts 28		X		X	X	X
Districts 29		X				X
Totals	7	15	1	4	3	15

Appendix F

Content Validity Index Survey Results 9/20/2012

9/20/2012	Not Relevant	Somewhat Relevant	Quite Relevant	Highly Relevant	Total	Percentage
4a	0	1	2	6	9	0.89
4b	0	0	0	9	9	1
4c	0	0	2	7	9	1
4d	0	2	2	5	9	0.78
5a	0	1	2	6	9	0.89
5b	0	0	2	7	9	1
5c	0	2	5	2	9	0.78
5d	3	2	2	2	9	0.44
5e	0	1	0	8	9	0.89
5f	0	0	2	7	9	1
6a	0	0	1	8	9	1
6b	0	0	2	7	9	1
6c	0	2	4	3	9	0.78
6d	0	0	0	9	9	1
6e	0	0	4	5	9	1
6f	0	1	1	7	9	0.89
6g	0	1	2	6	9	0.89
6h	0	0	1	8	9	1
7a	0	0	4	5	9	1
7b	0	2	4	3	9	0.78
7c	0	3	3	3	9	0.67
7d	0	1	2	6	9	0.89
7e	0	1	2	6	9	0.89
7f	0	1	4	4	9	0.89
7g	0	1	2	6	9	0.89
7h	0	1	1	7	9	0.89
7i	0	1	4	4	9	0.89
7j	0	0	6	3	9	1
7k	0	1	5	3	9	0.89
7l	0	1	5	3	9	0.89
7m	0	0	4	5	9	1
7n	0	3	2	4	9	0.67
7o	0	0	6	3	9	1
7p	1	4	1	3	9	0.44
8	0	0	5	4	9	1
9a	0	0	5	4	9	1

9b	1	0	3	5	9	0.89
9c	0	0	2	7	9	1
9d	0	0	2	7	9	1
10	0	0	5	4	9	1
11	0	0	3	6	9	1
12	0	1	3	5	9	0.89
13	0	1	2	6	9	0.89

Average 0.90

Appendix G

Content Validity Index Survey Results With Exclusions 9/20/2012

9/20/2012	Not Relevant	Somewhat Relevant	Quite Relevant	Highly Relevant	Total	Percentage
4a	0	1	2	6	9	0.89
4b	0	0	0	9	9	1
4c	0	0	2	7	9	1
4d	0	2	2	5	9	0.78
5a	0	1	2	6	9	0.89
5b	0	0	2	7	9	1
5c	0	2	5	2	9	0.78
5e	0	1	0	8	9	0.89
5f	0	0	2	7	9	1
6a	0	0	1	8	9	1
6b	0	0	2	7	9	1
6c	0	2	4	3	9	0.78
6d	0	0	0	9	9	1
6e	0	0	4	5	9	1
6f	0	1	1	7	9	0.89
6g	0	1	2	6	9	0.89
6h	0	0	1	8	9	1
7a	0	0	4	5	9	1
7b	0	2	4	3	9	0.78
7c	0	3	3	3	9	0.67
7d	0	1	2	6	9	0.89
7e	0	1	2	6	9	0.89
7f	0	1	4	4	9	0.89
7g	0	1	2	6	9	0.89
7h	0	1	1	7	9	0.89
7i	0	1	4	4	9	0.89
7j	0	0	6	3	9	1
7k	0	1	5	3	9	0.89
7l	0	1	5	3	9	0.89
7m	0	0	4	5	9	1
7n	0	3	2	4	9	0.67
7o	0	0	6	3	9	1
8	0	0	5	4	9	1
9a	0	0	5	4	9	1

9b	1	0	3	5	9	0.89
9c	0	0	2	7	9	1
9d	0	0	2	7	9	1
10	0	0	5	4	9	1
11	0	0	3	6	9	1
12	0	1	3	5	9	0.89
13	0	1	2	6	9	0.89

Average 0.92